

# THE GOOD, THE BAD AND THE UGLY: MULTIPLE ROLES OF BACTERIA IN HUMAN LIFE

EDITED BY: Tatiana Venkova, Chew Chieng Yeo and Manuel Espinosa  
PUBLISHED IN: *Frontiers in Microbiology*





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-574-4

DOI 10.3389/978-2-88945-574-4

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



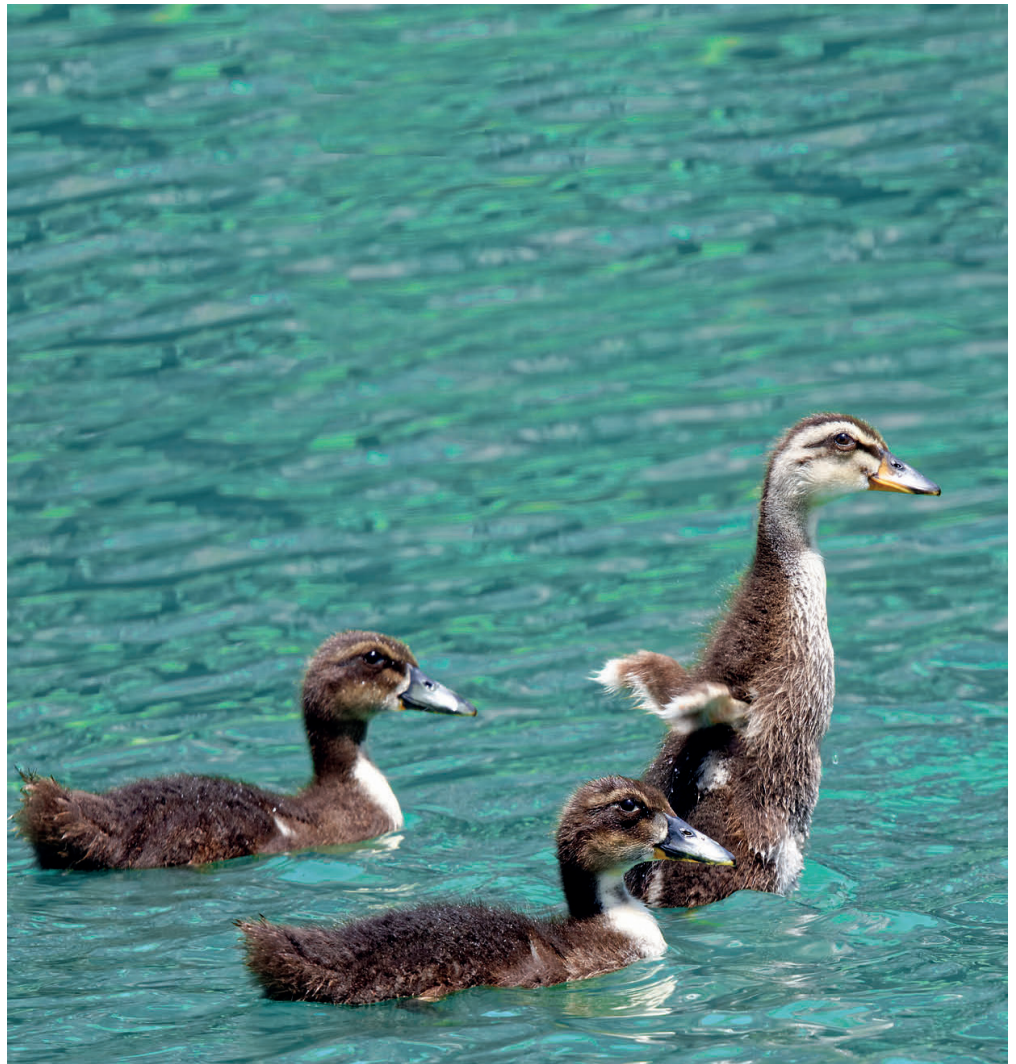
# THE GOOD, THE BAD AND THE UGLY: MULTIPLE ROLES OF BACTERIA IN HUMAN LIFE

Topic Editors:

**Tatiana Venkova**, Fox Chase Cancer Center, United States

**Chew Chieng Yeo**, University Sultan Zainal Abidin, Malaysia

**Manuel Espinosa**, Centro de Investigaciones Biológicas, Spain



"Los Tres Amigos". Image courtesy of Steve Kendall, [www.purplekitephoto.com](http://www.purplekitephoto.com)

Bacteria are among the earliest forms of life on Earth. Notwithstanding their small size and primitive origin, bacteria still have a tremendous impact on everyday human life. Over the centuries, research into bacteria have provided and enriched the fundamental biological knowledge due to their readily measured processes and effects on higher organisms. Although molecular genetics and microbiology were among the scientific fields that have mostly benefited from the discoveries made in bacteria, our current state of knowledge has gone beyond what anyone could have

ever imagined. The present Research Topic aims to cover new and exciting broad aspects of the importance of bacteria to human life, both positive and negative influences. Regulation of bacterial gene expression, replication and segregation control mechanisms, cell to cell communication via quorum sensors, and the relatively recent finding of bacterial immunity via CRISPR, have led to the development of many, and very important new tools in biotechnology and the emerging field of molecular medicine. The battle against infectious diseases has also benefited from the genetic approaches that have been developed in the quest for finding new targets and novel drugs against pathogenic bacteria. At the next level, the human microbiome project has opened up new avenues in understanding the role of bacteria in human health and wellbeing. Finally, the relationship between bacterial infections and human cancers will also be covered, a subject that is still under verification through rigorous experimental approaches. Special emphasis will be given to the bacterial accessory genome, i.e the mobilome, as the primary cause of health-threatening antimicrobial resistance and the production of toxins and virulence factors. Taking into account the evolutionary importance of horizontal gene transfer and the additional beneficial roles of certain bacterial mobile genetic elements, they help project best “the Good, the Bad and the Ugly” outline of this topic.

At the time this eBook is about to be published, our Research Topic has registered nearly 55,000 views.

**Citation:** Venkova, T., Yeo, C. C., Espinosa, M., eds (2018). The Good, The Bad and The Ugly: Multiple Roles of Bacteria in Human Life. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-574-4



# Table of Contents

## **08 Editorial: The Good, The Bad, and The Ugly: Multiple Roles of Bacteria in Human Life**

Tatiana Venkova, Chew Chieng Yeo and Manuel Espinosa

## **CHAPTER 1**

### **THE BENEFICIAL MICRO-WORLD**

#### **1.1. SOURCE OF FUNDAMENTAL BIOLOGY KNOWLEDGE**

#### **12 Scoring Targets of Transcription in Bacteria Rather Than Focusing on Individual Binding Sites**

Marko Djordjevic, Magdalena Djordjevic and Evgeny Zdobnov

#### **22 Successful Establishment of Plasmids R1 and pMV158 in a New Host Requires the Relief of the Transcriptional Repression of Their Essential rep Genes**

José Á. Ruiz-Masó, Luis M. Luengo, Inmaculada Moreno-Córdoba, Ramón Díaz-Orejas and Gloria del Solar

#### **40 The Importance of the Expendable: Toxin–Antitoxin Genes in Plasmids and Chromosomes**

Ramón Díaz-Orejas, Manuel Espinosa and Chew Chieng Yeo

#### **47 The Bacillus subtilis Conjugative Plasmid pLS20 Encodes Two Ribbon-Helix-Helix Type Auxiliary Relaxosome Proteins That are Essential for Conjugation**

Andrés Miguel-Arribas, Jian-An Hao, Juan R. Luque-Ortega, Gayetri Ramachandran, Jorge Val-Calvo, César Gago-Córdoba, Daniel González-Álvarez, David Abia, Carlos Alfonso, Ling J. Wu and Wilfried J. J. Meijer

#### **59 Bad Phages in Good Bacteria: Role of the Mysterious orf63 of $\lambda$ and Shiga Toxin-Converting $\Phi 24_{\text{B}}$ Bacteriophages**

Aleksandra Dydecka, Sylwia Bloch, Ali Rizvi, Shaili Perez, Bożena Nejman-Falenczyk, Gracja Topka, Tomasz Gasior, Agnieszka Necel, Grzegorz Wegrzyn, Logan W. Donaldson and Alicja Wegrzyn

#### **71 The Transcriptome of Streptococcus pneumoniae Induced by Local and Global Changes in Supercoiling**

Adela G. de la Campa, María J. Ferrándiz, Antonio J. Martín-Galiano, María T. García and Jose M. Tirado-Vélez

#### **89 Toxin $\zeta$ Triggers a Survival Response to Cope With Stress and Persistence**

María Moreno-del Álamo, Mariangela Tabone, Virginia S. Lioy and Juan C. Alonso

#### **1.2. USE IN BIOTECHNOLOGY**

#### **100 Molecular Basis of Stationary Phase Survival and Applications**

Jananee Jaishankar and Preeti Srivastava

#### **112 Features of CRISPR-Cas Regulation Key to Highly Efficient and Temporally-Specific crRNA Production**

Andjela Rodic, Bojana Blagojevic, Magdalena Djordjevic, Konstantin Severinov and Marko Djordjevic

### 1.3. PROBIOTICS

- 128** *Ribonucleotide Reductases From Bifidobacteria Contain Multiple Conserved Indels Distinguishing Them From all Other Organisms: In Silico Analysis of the Possible Role of a 43 aa Bifidobacteria-Specific Insert in the Class III RNR Homolog*  
Seema Alnajar, Bijendra Khadka and Radhey S. Gupta
- 142** *Characterization of the Sorbitol Utilization Cluster of the Probiotic *Pediococcus parvulus* 2.6: Genetic, Functional and Complementation Studies in Heterologous Hosts*  
Adrian Pérez-Ramos, Maria L. Werning, Alicia Prieto, Pasquale Russo, Giuseppe Spano, Mari L. Mohedano and Paloma López
- 159** *Dextranucrase Expression is Concomitant With That of Replication and Maintenance Functions of the *pMN1* Plasmid in *Lactobacillus sakei* MN1*  
Montserrat Nàcher-Vázquez, José A. Ruiz-Masó, María L. Mohedano, Gloria del Solar, Rosa Aznar and Paloma López

### 1.4. ENVIRONMENTAL BIOREMEDIATION

- 175** *Plasmid-Mediated Bioaugmentation for the Bioremediation of Contaminated Soils*  
Carlos Garbisu, Olatz Garaiurrebaso, Lur Epelde, Elisabeth Grohmann and Itziar Alkorta
- 188** *Comparative Genomic Analysis Reveals Organization, Function and Evolution of *ars* Genes in *Pantoea* spp.*  
Liyang Wang, Jin Wang and Chuanyong Jing

## CHAPTER 2

### PATHOGENS UNVEILED

#### 2.1. MOLECULAR MECHANISMS

- 200** *Plasmid Replicons From *Pseudomonas* are Natural Chimeras of Functional, Exchangeable Modules*  
Leire Bardaji, Maite Añorga, José A. Ruiz-Masó, Gloria del Solar and Jesús Murillo
- 214** *Small, Enigmatic Plasmids of the Nosocomial Pathogen, *Acinetobacter baumannii*: Good, Bad, Who Knows?*  
Soo Sum Lean and Chew Chieng Yeo
- 222** **Fic* Proteins of *Campylobacter fetus* Subsp. *Venerealis* Form a Network of Functional Toxin–Antitoxin Systems*  
Hanna Sprenger, Sabine Kienesberger, Brigitte Pertschy, Lisa Pörtl, Bettina Konrad, Priya Bhutada, Dina Vorkapic, Denise Atzmüller, Florian Feist, Christoph Högenauer, Gregor Gorkiewicz and Ellen L. Zechner
- 239** *A Disulfide Bond in the Membrane Protein *IgaA* Is Essential for Repression of the *RcsCDB* System*  
M. Graciela Pucciarelli, Leticia Rodríguez and Francisco García-del Portillo
- 249** *Fluorescence Imaging of *Streptococcus pneumoniae* With the Helix pomatia Agglutinin (HPA) as a Potential, Rapid Diagnostic Tool*  
Mirian Domenech and Ernesto García



## 2.2. IN-DEPTH ANTIBIOTIC RESISTANCE

### 261 ***Molecular Mechanisms That Contribute to Horizontal Transfer of Plasmids by the Bacteriophage SPP1***

Ana Valero-Rello, María López-Sanz, Alvaro Quevedo-Olmos, Alexei Sorokin and Silvia Ayora

### 274 ***Phages in the Human Body***

Ferran Navarro and Maite Muniesa

### 281 ***Replication of Staphylococcal Resistance Plasmids***

Stephen M. Kwong, Joshua P. Ramsay, Slade O. Jensen and Neville Firth

### 297 ***Biofilm-Forming Clinical Staphylococcus Isolates Harbor Horizontal Transfer and Antibiotic Resistance Genes***

Sandra Águila-Arcos, Itxaso Álvarez-Rodríguez, Olatz Garaiurrebaso, Carlos Garbisu, Elisabeth Grohmann and Itziar Alkorta

### 309 ***PCR-Based Analysis of ColE1 Plasmids in Clinical Isolates and Metagenomic Samples Reveals Their Importance as Gene Capture Platforms***

Manuel Ares-Arroyo, Cristina Bernabe-Balas, Alfonso Santos-Lopez, Maria R. Baquero, Kashi N. Prasad, Dolores Cid, Carmen Martin-Espada, Alvaro San Millan and Bruno Gonzalez-Zorn

### 324 ***Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal Salmonella enterica***

Saskia Neuert, Satheesh Nair, Martin R. Day, Michel Doumith, Philip M. Ashton, Kate C. Mellor, Claire Jenkins, Katie L. Hopkins, Neil Woodford, Elizabeth de Pinna, Gauri Godbole and Timothy J. Dallman

### 335 ***Antibiotic-Induced Alterations in Gut Microbiota are Associated With Changes in Glucose Metabolism in Healthy Mice***

Richard R. Rodrigues, Renee L. Greer, Xiaoxi Dong, Karen N. DSouza, Manoj Gurung, Jia Y. Wu, Andrey Morgun and Natalia Shulzhenko

## 2.3. GENOMICS AND EVOLUTION OF PATHOGENIC BACTERIA

### 349 ***The Intriguing Evolutionary Journey of Enteroinvasive E. coli (EIEC) Toward Pathogenicity***

Martina Pasqua, Valeria Michelacci, Maria Letizia Di Martino, Rosangela Tozzoli, Milena Grossi, Bianca Colonna, Stefano Morabito and Gianni Prosseda

### 361 ***Environmental Origin of the Genus Bordetella***

Illiassou Hamidou Soumana, Bodo Linz and Eric T. Harvill

### 371 ***Mycobacterium tuberculosis Acquires Limited Genetic Diversity in Prolonged Infections, Reactivations and Transmissions Involving Multiple Hosts***

Marta Herranz, Ilva Pole, Iveta Ozere, Álvaro Chiner-Oms, Miguel Martínez-Lirola, Felipe Pérez-García, Paloma Gijón, María Jesús Ruiz Serrano, Laura Clotet Romero, Oscar Cuevas, Iñaki Comas, Emilio Bouza, Laura Pérez-Lago and Darío García-de-Viedma

### 380 ***In-Depth Characterization and Functional Analysis of Clonal Variants in a Mycobacterium tuberculosis Strain Prone to Microevolution***

Yurena Navarro, Laura Pérez-Lago, Marta Herranz, Olalla Sierra, Iñaki Comas, Javier Sicilia, Emilio Bouza and Darío García de Viedma

**388 *Double-Face Meets the Bacterial World: The Opportunistic Pathogen *Stenotrophomonas maltophilia****

Felipe Lira, Gabriele Berg and José L. Martínez

**2.4. INFECTIOUS DISEASES AND LINK TO CANCER**

**403 *Secondary Bacterial Infections Associated With Influenza Pandemics***

Denise E. Morris, David W. Cleary and Stuart C. Clarke

**420 *The Pivotal Role of DNA Repair in Infection Mediated-Inflammation and Cancer***

Ayşe Z. Sahan, Tapas K. Hazra and Soumita Das

**CHAPTER 3**

**BACTERIA AND HUMAN LIFE**

**3.1. NOVEL ANTIBIOTICS**

**437 *Targeting the Bacterial Orisome in the Search for New Antibiotics***

Julia E. Grimwade and Alan C. Leonard

**444 *Conjugation Inhibitors and Their Potential Use to Prevent Dissemination of Antibiotic Resistance Genes in Bacteria***

Elena Cabezón, Fernando de la Cruz and Ignacio Arechaga

**451 *A Novel Antimicrobial Coating Represses Biofilm and Virulence-Related Genes in Methicillin-Resistant *Staphylococcus aureus****

Ankita Vaishampayan, Anne de Jong, Darren J. Wight, Jan Kok and Elisabeth Grohmann

**3.2. DRUG DELIVERY AND CANCER THERAPY**

**465 *DNA Delivery and Genomic Integration into Mammalian Target Cells Through Type IV A and B Secretion Systems of Human Pathogens***

Dolores L. Guzmán-Herrador, Samuel Steiner, Anabel Alperi, Coral González-Prieto, Craig R. Roy and Matxalen Llosa

**477 *Bacterial Therapy of Cancer: Promises, Limitations, and Insights for Future Directions***

M. Gabriela Kramer, Martín Masner, Fernando A. Ferreira and Robert M. Hoffman

**3.3. NEW ASPECTS**

**486 *Outlining Core Pathways of Amyloid Toxicity in Bacteria With the RepA-WH1 Prionoid***

Laura Molina-García, María Moreno-del Álamo, Pedro Botias, Zaira Martín-Moldes, María Fernández, Alicia Sánchez-Gorostiaga, Aída Alonso-del Valle, Juan Nogales, Jesús García-Cantalejo and Rafael Giraldo

**507 *Cadaver Thanatomicrobiome Signatures: The Ubiquitous Nature of *Clostridium* Species in Human Decomposition***

Gulnaz T. Javan, Sheree J. Finley, Tasia Smith, Joselyn Miller and Jeremy E. Wilkinson





# Editorial: The Good, The Bad, and The Ugly: Multiple Roles of Bacteria in Human Life

Tatiana Venkova<sup>1\*</sup>, Chew Chieng Yeo<sup>2</sup> and Manuel Espinosa<sup>3</sup>

<sup>1</sup> Fox Chase Cancer Center, Research & Development Alliances, Rockledge, PA, United States, <sup>2</sup> Faculty of Medicine, University Sultan Zainal Abidin Medical Campus, Kuala Terengganu, Malaysia, <sup>3</sup> Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

**Keywords:** infectious diseases, antimicrobial resistance, virulence, bacterial immunity, horizontal gene spread, mobile genetic elements, bacteria and cancer

## Editorial on the Research Topic

### The Good, The Bad, and The Ugly: Multiple Roles of Bacteria in Human Life

## OPEN ACCESS

### Edited by:

Marina G. Kalyuzhanaya,  
San Diego State University,  
United States

### Reviewed by:

Miguel Angel Cevallos,  
Universidad Nacional Autónoma de  
México, Mexico  
Dhruba Chattoraj,  
National Institutes of Health (NIH),  
United States  
Antonius Suwanto,  
Bogor Agricultural University,  
Indonesia

### \*Correspondence:

Tatiana Venkova  
venkova@hotmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 07 May 2018

**Accepted:** 09 July 2018

**Published:** 27 July 2018

### Citation:

Venkova T, Yeo CC and Espinosa M  
(2018) Editorial: The Good, The Bad,  
and The Ugly: Multiple Roles of  
Bacteria in Human Life.  
Front. Microbiol. 9:1702.  
doi: 10.3389/fmicb.2018.01702

## INTRODUCTION

*"If you don't like bacteria, you are on the wrong planet."*

(Brand, 2010).

Quoting the writer and editor Stewart Brand, summarizes the solid facts, knowledge, and fascination that we all share with regard to the smallest and simplest organisms on Earth. Bacteria are not only considered the cradle of Life, but as revealed by history and centuries of scientific interest, they are the living organisms that affect us, the Humans, most. From the moment Antonie van Leeuwenhoek observed for the first time the tiny bacterial cells under the microscope, up until the ongoing sequencing projects on the human microbiome, it has been and is an exciting journey of understanding, fighting, and using bacteria for our benefit. Many a time we tend to anthropomorphise our subjects of study, which is not necessarily a wrong practice if we remain aware of our doings and of our conclusions, thus we can artificially classify bacteria into "beneficial or pathogenic" in unequal proportion. However, with the knowledge gained throughout the years, we are still under the impression that it is still enigmatic whether we can consider bacteria as "The Good" or "The Bad" and "The Ugly" that co-habits with us. This is precisely what we have tried to do under this Research Topic with such a well-known and anthropological name, in which we have tried to combine different aspects of the bacterial world and to show how bacteria strongly influence our lifestyle. Of course, we are aware that drawing lines is a risky exercise, because what to do when a "Good" converts itself into a "Bad" and "Ugly"? Enterococci are a good example: from being a respectable member of our gut microbiome, it can turn Ugly given certain circumstances (low immuno-response on their host, we, Humans). Being scientists and trying to guide the present Research Topic, we take the scientific approach in addressing such a complex and difficult task by presenting facts and drawing conclusions that should help the readers appreciate the fascinating full spectrum of the roles that bacteria play in human life.

First and the foremost, Molecular Biology would not be where it stands today were it not for the knowledge of the basic blocks of life, i.e. DNA, RNA, and proteins, and of processes such as gene expression and its control, chromosome replication and cell division, horizontal gene transfer, cell to cell communication, DNA repair, cell immunity and cell death, that were obtained from studies in bacteria. As all these processes are relatively easy to measure in bacteria, and that the basic principles of biological regulation being same in all organisms, the knowledge gained in studying bacteria is benefiting biological sciences as a whole, including biotechnology and the emerging

field of molecular medicine. In these, plasmids and phages, i.e. the bacterial mobilome, play an important role. Plasmids and phages were not only the main platform of the fundamental biological discoveries, but they are also a versatile tool for gene delivery in all organisms and the main reason for the spread of antibiotic resistance that takes a harsh toll on human life and the economy. Because many bacteria that are opportunistic pathogens live in symbiosis with plants or inhabit polluted environments, carry plasmids with genes for resistance or production of a particular enzyme, the scientific community is in a quest for finding new antibiotics to deal with infections caused by pathogenic bacteria, increase yield of plants and stimulate biodegradation. Bacterial plasmids have also been associated with and considered the “culprit” for beneficial production of animal and human food and beverages. Probiotics, which stimulate immunity and anti-inflammation, and the increasing reports of bacteria linked to cancer, are the two ultimate examples of the opposite Good and Ugly sides of bacteria. Until recently, we knew only that bacteria inhabit soil, water, extreme environments such as acidic hot springs and radioactive waste, and live in symbiotic and parasitic relationships with plants and animals. The human microbiome project opened new avenues in our understanding on the close relationship between bacteria and humans. The fact that our body, on the inside and its surface, is heavily inhabited by bacteria, urges the need for deeper investigation and we are hopeful the current Topic will provide new clues and valuable information supporting the importance of studying bacteria.

## THE RESEARCH TOPIC

Despite the vast information available to date and the general belief that bacteria are more harmful than beneficial to the human population, the mere intent of proposing this Research Topic was to probe the current state of knowledge on bacteria and to figure out whether they affect our life simply in a negative/positive way, or the picture is more complex than we could have imagined. We were delighted to see that the Topic attracted the attention of 214 authors from 5 continents that responded enthusiastically with 40 original research and review articles. Our colleagues were from different scientific fields with diverse interests and points of views, which enriched enormously our understanding and knowledge on the subject. It is thus our pleasure to present the contributors to the Research Topic “The Good, The Bad and The Ugly: Multiple Roles of Bacteria in Human Life” with their invaluable reports.

## CHAPTERS

### Chapter 1: The Beneficial Micro-World

- 1.1. Source of Fundamental Biology Knowledge
- 1.2. Use in Biotechnology
- 1.3. Probiotics
- 1.4. Environmental Bioremediation

### Chapter 2: Pathogens unveiled

- 2.1. Molecular Mechanisms

- 2.2. In-Depth Antibiotic Resistance
- 2.3. Genomics and Evolution of pathogenic bacteria
- 2.4. Infectious Diseases and Link to Cancer

## Chapter 3: Bacteria and Human Life

- 3.1. Novel Antibiotics
- 3.2. Drug Delivery and Cancer Therapy
- 3.3. New Aspects

The Topic initiates with papers reporting on the fundamental biological discoveries that enable deeper understanding of bacterial gene expression and draw more accurate models for predicting bacterial transcription targets (Djordjevic et al.), the circuits of regulation of the replication genes in bacterial plasmids for their successful establishment in new hosts (Ruiz-Masó et al.), and a review on the possible biological roles of type II toxin-antitoxin modules, in both plasmids and chromosomes, showing evidence of the functional overlap of these modules irrespective of their genomic location (Díaz-Orejas et al.). The following two articles then deal with fundamental discoveries in the mobilome, such as the finding of two previously unknown proteins participating in the mobilization complex (relaxosome) encoded by plasmid pLS20 (Miguel-Arribas et al.), and the characterization of a mysterious protein encoded by lambda and lambdoid phages (Dydecka et al.) that may play an important role in the regulation of the decision of these phages in becoming “Ugly” (lysogeny) or “Real Bad” (lytic) in this case for the bacterial host. Two more articles relate to the Firmicutes lifestyle: how to deal with the chromosomal supercoils and the expression of genes in the pneumococcus (De La Campa et al.), one of the “Bad Ones,” usually acting as a harmless commensal in our nasopharynx, but ready to strike pneumonia when our immune system goes down, and the second dealing with the response of bacteria to stressful situations that lead to another decision: to be swept away by the stress or to survive in a dormant persister state, thus permitting the bacteria to cope with adverse (for the bacterium) situations, like facing antibiotic treatments (Moreno-del Álamo et al.).

The uses of bacteria in Biotechnology is covered by an extensive review on bacterial stationary phase promoters and their application for construction of improved gene-expression systems in recombinant protein production and in the bioremediation processes (Jaishankar and Srivastava). The “hot” topic of the CRISPR-Cas bacterial immune system that has been famously utilized in the gene-editing of mammalian cells in recent years is tackled by a closer look into its fine-tuned regulation and the proposed efficient expression of small RNAs in a narrow time interval (Rodic et al.). New insights on the genotype, enzyme production and physiological properties of beneficial bacteria such as the well-known probiotic *Bifidobacteria* (Alnajar et al.) and the newly-described *Pediococcus parvulus* (Pérez-Ramos et al.) are presented in depth, with a special emphasis on the role of their plasmids as in *Lactobacillus sakei* (Nácher-Vázquez et al.). Plasmids are also the main mediator of bioremediation of contaminated soils as reported by two independent groups (Garbisu et al.; Wang et al.).



The theme of the bacterial mobile genetic elements, plasmids, and phages, is extensively covered in our next Chapter (Pathogens Unveiled), as they are the driving force for horizontal gene transfer and the main cause of antibiotic resistance and virulence. We learned the interesting fact that the virulence of the pathogen *Pseudomonas syringae* is mediated by natural chimeras of distinct plasmid families (Bardaji et al.). Lean and Yeo examine our current knowledge of plasmids that are less than 10 kb in size commonly found in the nosocomial pathogen, *Acinetobacter baumannii*, in a mini-review. Some of these small plasmids harbor resistance as well as potential virulence genes whereas others are truly enigmatic. An interesting article relates the wide-spread world of prokaryotic toxin-antitoxin systems to bacterial virulence in the important pathogen of the *Campylobacter* genus, one of the “Bad Ones” because of their multiple resistances to antibiotics and their clinical relevance. Sprenger et al. show that in *Campylobacter fetus* subspecies *venerealis*, the activity of some FIC (filamentation induced by cyclic AMP) proteins resemble classical TA systems and appeared to be related to virulence. Graciela Pucciarelli et al. examine in detail the role of a disulfide bond in the major periplasmic loop of the IgaA inner membrane protein of another pathogen, *Salmonella enterica* serovar Typhimurium, in the regulation of the RcsCDB phosphorelay system, which is involved in regulating the expression of a multitude of cellular processes including motility, biofilm production and virulence. The potential use of the lectin produced by the edible snail, *Helix pomatia* agglutinin (HPA) as a novel diagnostic tool for the identification of *Streptococcus pneumoniae* is proposed by Domenech and Garcia who show that the HPA lectin specifically recognizes the terminal  $\alpha$ GalNAc residues of the cell wall teichoic and lipoteichoic acids of *S. pneumoniae*.

The role of the bacterial viruses, the bacteriophages (or just phages), in the rapid dissemination of antibiotic resistance is presented by Valero Rello et al., whereas the entire spectrum of their impact on human health is summarized in the review article by Navarro and Muniesa. It is important to remember that bacterial phages played (and still do!) a key role in the early stages of Molecular Biology research, since they enabled the study of the control of gene expression and decision-making responses, which led to the development of controlled expression systems for protein over-expression. Further, the number of bacteriophages on planet Earth (around  $10^{31}$ ) is more than any other organism, including bacteria, combined, making them a formidable evolutionary driving force.

Plasmids as vehicles for horizontal (lateral) transfer of antibiotic-resistance traits and their “evil” doings are represented by important contributions in both the Gram-positive and the Gram-negative bacterial pathogens. Identification of these genetic elements and the ways they perform their role in gene transfer are major problems nowadays, when the number of new antibacterials are dwindling. An excellent review on the replication mechanisms of several staphylococcal plasmids that mediate antimicrobial resistance is presented by Kwong et al. Águila-Arcos et al. show that in all 25 biofilm-forming clinical staphylococcal isolates that were studied, horizontal transfer and relaxase genes of two common staphylococcal

resistance plasmids, pSK41 and pT181, were detected, inferring the possibility of the dissemination of antibiotic resistance to other clinical isolates. In another paper, Ares-Arroyo et al. analyze various ColE1 replicons using bioinformatics and experimental approaches. They developed a new PCR-based system for the detection and analysis of ColE1 plasmids and validated their important role in the dissemination of antibiotic resistance. Whole genome sequencing (WGS) has been routinely implemented for the identification and surveillance of *Salmonella* at Public Health England’s Gastrointestinal Bacteria Reference Unit since 2014. Neuert et al. evaluated the prediction of phenotypic antimicrobial resistance in non-typhoidal *Salmonella enterica* from the genotypic profiles obtained from the whole genome sequences of 3,491 isolates received between 2014 and 2015 by Public Health England and showed that discrepancies between phenotypic and genotypic profiles were low and that by and large, WGS is suitable as a rapid means of determining antimicrobial resistance profiles for surveillance.

Looking at the other side of the coin, the study on gut microbiota and the changes in gene expression and glucose metabolism induced by antibiotic treatment shows the complex nature of our choices onto how to fight bad bacteria (Rodrigues et al.).

Taking into account the medical and environmental impact of bacterial pathogens, a special emphasis is given on understanding their genomics and evolution. *Escherichia coli* and *Bordetella*, two of the most devastating human and animals pathogens are covered extensively (Pasqua et al. and Hamidou Soumana et al.). A very important example of our change of views from “forgotten” to “Real Bad” bacteria is provided by the tuberculosis pathogen. Two papers show that *Mycobacterium tuberculosis* causing prolonged infections can acquire a limited genetic diversity, and yet there are strains prone to microevolution within the infected host (Herranz et al.; Navarro et al.). Lira et al. present comparative genomic analyses of the opportunistic pathogen, *Stenotrophomonas maltophilia*, obtained from clinical as well as environmental samples and show that there are no distinct or separate clinical and environmental lineages of the pathogen. This indicates that infection is mainly due to impaired immune response of infected patients and given the biotechnological potential of *S. maltophilia*, its use in its natural habitats will likely only lead to an incremental risk in acquiring infections.

Morris et al. present a comprehensive review on the role of secondary bacterial infections in increasing the morbidity and mortality of influenza infections, especially during epidemics and pandemics. The increasing antimicrobial resistance and vaccine evasion presented by these bacterial pathogens have made it even more crucial to monitor their epidemiology to better guide clinical treatment and development particularly during an influenza epidemic or pandemic. In another review, Sahan et al. show how pathogenic microorganisms can induce various levels of inflammation which can lead to DNA damage, thereby posing a risk for the development of cancers. The review focuses on *Helicobacter pylori*-mediated inflammation and gastric cancer as well as the potential role of *Fusobacterium*

*nucleatum* in colorectal cancers besides indicating the important role of DNA repair pathways in precluding the development of such cancers.

The last Chapter (Bacteria and Human Life) starts with tackling the current shortage of effective treatment for bacterial infections and the quest for new antibiotics, being a priority of the scientific microbiological community. In a Perspective article, Grimwade and Leonard examine our current knowledge regarding the initiation stage of bacterial chromosomal replication, mediated by the bacterial ori<sub>se</sub>, and they identify potential targets that could prevent bacterial chromosomal replication, which therefore could serve as targets for novel antibacterial compounds.

Molecules that are able to inhibit conjugation (COINS) have been proposed and thought to provide a novel avenue to combat the spread of antibiotic-resistance traits encoded by mobile elements. Some effective COINS were discovered a few years ago, and the strategies to identify and to further develop them are reviewed by Cabezón et al. Another approach to deal with pathogens is related to the hindering of their lifestyle. Many bacteria (and the Bad Ones are no exception) grow happier when they grow together forming biofilms that will stick to living (teeth, nasopharynx) or implant (catheters, prosthesis) surfaces to better communicate among them as well as to colonize new niches. Inhibitors of biofilm formation are an interesting source of potential antimicrobial drugs that is recorded by Vaishampayan et al.

One of the major difficulties in the drug discovery field is how to deliver the desired drug so that it reaches its final target. Drug delivery strategies can be costly and time-consuming; this could make a cleverly designed drug unable to be used because of the lack of proper delivery procedures. Exploring this field has been the subject of the Llosa's laboratory for years. They provide now an insightful review on how to use the bacterial Type IV secretion system pathways to deliver and to stably integrate into mammalian cell chromosomes with desired traits that would, in time, lead to anti-tumor drugs (Guzmán-Herrador et al.).

The use of several bacterial species in cancer therapy is examined in a Perspective article by Gabriela Kramer et al., thus contrasting with the “Bad” side of other bacterial species that have been shown to be potential causal agents for cancer (Sahan et al.).

## REFERENCES

Brand S. (2010). *Whole Earth Discipline: Why Dense Cities, Nuclear Power, Transgenic Crops, Restored Wildlands, and Geoengineering Are Necessary*. New York, NY: Penguin Publishing Group.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Intriguingly, attenuated mutants of the pathogen, *Salmonella enterica* serovar Typhimurium have been shown to invade and destroy a broad range of cancer cell types *in vitro* and are so far, the most efficient anti-tumor bacteria in experimental models of cancer.

Finally, unexpected facts on bacteria are unveiled by the last 2 articles in our Topic. Molina-García et al. show that bacteria can be the ideal model for studying human neurodegenerative diseases, whereas Javan et al. report on the bacterial thanatomicrobiome that could aid in forensic investigations.

## WHAT IS NEXT?

Despite the vast knowledge on bacteria, including the current Research Topic, new and exciting scientific reports are coming up every day. Among those, the human microbiome project plays a central role on revealing the true interaction between us, the humans, and those “primitive” but powerful living organisms that have now been shown to play central roles in shaping our health and our environment. As the collection of articles in this Research Topic has shown, bacteria do indeed display all facets of the “Good,” the “Bad,” and the “Ugly,” and like everything else in this world of ours, all three facets co-exist as a dynamic, chaotic whole.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

While this Editorial was written, authors were funded by FRGS/1/2016/SKK11/UNISZA/01/1 (CCY) and by BIO2015-69085-REDC (ME).

## ACKNOWLEDGMENTS

We warmly thank all the contributors to this eBook, the dedicated reviewers, and the editorial support of the Journal.

Copyright © 2018 Venkova, Yeo and Espinosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Scoring Targets of Transcription in Bacteria Rather than Focusing on Individual Binding Sites

Marko Djordjevic<sup>1\*</sup>, Magdalena Djordjevic<sup>2</sup> and Evgeny Zdobnov<sup>3</sup>

<sup>1</sup> Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Belgrade, Serbia, <sup>2</sup> Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia, <sup>3</sup> Swiss Institute of Bioinformatics and Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Alexandre V. Morozov,  
Rutgers University, The State  
University of New Jersey,  
United States  
Yuriy L. Orlov,  
Institute of Cytology and Genetics  
(RAS), Russia  
Anastasia Anashkina,  
Engelhardt Institute of Molecular  
Biology (RAS), Russia

### \*Correspondence:

Marko Djordjevic  
dmarko@bio.bg.ac.rs

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 31 July 2017

**Accepted:** 09 November 2017

**Published:** 22 November 2017

### Citation:

Djordjevic M, Djordjevic M and  
Zdobnov E (2017) Scoring Targets  
of Transcription in Bacteria Rather  
than Focusing on Individual Binding  
Sites. *Front. Microbiol.* 8:2314.  
doi: 10.3389/fmicb.2017.02314

Reliable identification of targets of bacterial regulators is necessary to understand bacterial gene expression regulation. These targets are commonly predicted by searching for high-scoring binding sites in the upstream genomic regions, which typically leads to a large number of false positives. In contrast to the common approach, here we propose a novel concept, where overrepresentation of the scoring distribution that corresponds to the entire searched region is assessed, as opposed to predicting individual binding sites. We explore two implementations of this concept, based on Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) tests, which both provide straightforward *P*-value estimates for predicted targets. This approach is implemented for pleiotropic bacterial regulators, including  $\sigma^{70}$  (bacterial housekeeping  $\sigma$  factor) target predictions, which is a classical bioinformatics problem characterized by low specificity. We show that KS based approach is both faster and more accurate, departing from the current paradigm of AD being slower, but more accurate. Moreover, KS approach leads to a significant increase in the search accuracy compared to the standard approach, while at the same time straightforwardly assigning well established *P*-values to each potential target. Consequently, the new KS based method proposed here, which assigns *P*-values to fixed length upstream regions, provides a fast and accurate approach for predicting bacterial transcription targets.

**Keywords:** direct target gene predictions, transcription factor binding site predictions, transcription regulation, position specific weight matrices, transcription targets, transcription start starts, sigma70, bacterial gene expression regulation

## INTRODUCTION

Identifying targets of transcription regulators (transcription targets), such as genes that are directly regulated by a given transcription factor, or transcribed by a certain  $\sigma$  factor, is a crucial step toward understanding bacterial gene expression regulation. Such knowledge is in turn crucial for both biotechnology applications and fundamental understanding of how bacteria respond to changing environment (e.g., during host pathogen interactions).

The task of identifying transcription targets is typically exhibited by starting from either: (i) large scale *in vivo* binding experiments such as ChIP-Seq (Wade et al., 2007; Park, 2009), (ii) large scale *in vitro* binding data, such as high-throughput SELEX (Roulet et al., 2002; Jagannathan et al., 2006) and protein binding microarrays (PBM) (Bulyk, 2006; Newburger and Bulyk, 2009),

and (iii) smaller scale experiments, such as SELEX, primer extension (for  $\sigma$  factors) and DNA footprinting (Green et al., 1989; Tuerk and Gold, 1990), which are typically assembled in databases such as TRANSFAC (Wingender, 2008), JASPAR (Mathelier et al., 2016), or RegulonDB (Gama-Castro et al., 2008). From these binding experiments, specificity of a given transcription factor (TF) is then extracted through some of the numerous methods that have been developed for this purpose. Those methods can be based on either information theory considerations (Stormo, 2000; Bulyk, 2004; Favorov et al., 2005; Ozoline and Deev, 2006; Levitsky et al., 2014; Korostelev et al., 2016), or on biophysical models (Stormo and Fields, 1998; Djordjevic et al., 2003; Djordjevic and Sengupta, 2006; Stormo and Zhao, 2010; Vilar, 2010; Djordjevic, 2013; Vilar and Saiz, 2013; Locke and Morozov, 2015), but in either case the inferred DNA binding specificity is represented in a form of a matrix, often called position specific weight matrix (PSWM). Note that, in the case of biophysics based approaches, these PSWMs in fact correspond to the so-called energy matrix (Djordjevic et al., 2003; Stormo and Zhao, 2010). These methods, up to now, have been shown to be able to extract the binding specificity with a reasonable accuracy, particularly when the data are coming from (controlled) high-throughput *in vitro* experiments (Bulyk, 2004, 2006; Djordjevic and Sengupta, 2006).

Once PSWMs are inferred, in prokaryotes they are used to scan genomic regions upstream of potential targets (e.g., the upstream intergenic regions), to find putative direct regulatory targets (Kim and Ren, 2006). These putative targets are next typically compared with the results of high-throughput experiments, such as DNA microarray data, or crosschecked with results of *in vivo* binding experiments (e.g., with the locations of binding peaks from ChIP-Seq experiments). This crosschecking may provide comprehensive information on the underlying regulatory mechanism, e.g., to what extent binding of the regulator under the given experimental conditions matches with the putative list of the genomic regions to which it is expected to bind. Such information is particularly useful when the binding specificity is inferred from *in vitro* binding studies, and is then crosschecked with independent experiments coming from *in vivo* binding measurements (Kim and Ren, 2006; Stormo and Zhao, 2010).

Despite the importance of accurately predicting direct targets for a given regulator, the bulk of the research efforts concentrate on more accurately inferring PSWM. On the other hand, a typical procedure for identifying putative direct targets in bacteria is rather simple, and involves scanning the upstream genomic regions by the inferred PSWM (Kim and Ren, 2006; Wade et al., 2007; Stormo and Zhao, 2010; de Jong et al., 2012). The sites with maximal PSWM scores are then identified, and those above certain thresholds are classified as putative targets. This procedure, however, often results in low search accuracy, in particular, in a very large number of false positives (Robison et al., 1998; Stormo, 2000). In eukaryotes, methods that predict clusters of transcription factor binding sites (TFBS) are also used, in addition to predicting individual TFBS. However, to successfully apply these methods, one often has to know which TFs functionally interact (Hannenhalli, 2008). Also, a recent

evaluation shows that the clustering methods lead to lower accuracy compared to individual TFBS predictions (Jayaram et al., 2016). The major reason behind the apparent low accuracy in the search of direct target genes is that individual high-scoring binding sites can easily appear by random chance in a sufficiently long genomic sequence, leading to so called non-sites (Kim and Ren, 2006). While this problem may be, to some extent, alleviated by negative selection acting on these non-sites, this negative selection is likely small. Furthermore, another problem, accurately assigning statistical significance to the targets predicted in such approach is also not well explored. That is, the maximal scoring sites are located in the tale of the weight matrix score distribution, and accurately calculating this tale requires doing an inverse Laplace transform of the corresponding partition function, which, in itself, is an ill-resolved numerical problem (Hertz and Stormo, 1999). Consequently, putative targets above certain threshold are typically reported without assigning statistical significance to the corresponding hits.

To address the problem of accurate transcription target predictions, we here develop a new concept which is based on the following hypothesis. We propose that, rather than identifying individual sites with high weight matrix scores, a better measure is assessing enrichment of the high scoring sites over a certain background in the entire region that is searched. This proposal then does not depend on individual high-scoring sites (which can easily emerge by random), but instead on comparing the weight matrix score distribution for the entire searched region with a certain background distribution. Note that this automatically accounts for the random occurrence of high-scoring binding sites, since such random occurrences (non-sites) would also appear in the background distribution. Moreover, this hypothesis directly couples with elegant statistical methods that allow determining statistical significance of a difference between the two distributions, such as Kolmogorov–Smirnov (KS) or Anderson–Darling (AD) tests. Therefore, these statistical tests also allow straightforwardly assigning a well-established statistical significance to the predicted direct targets, which also addresses the other major deficiency of the usual approach discussed above. Consequently, in contrast to the previous approaches, we will here develop a method which is based on assigning P values to fixed length upstream regions (e.g., the upstream intergenic regions in bacteria), rather than picking up only the best scoring PSWM matches (or their clusters).

However, significant questions emerge with regard to our proposed novel concept:

- (i) Can search based on this hypothesis indeed identify direct targets with high accuracy? Does the classification threshold, based on statistical significance assigned through this approach, lead to high prediction accuracy?
- (ii) What is the appropriate background (null distribution)?
- (iii) What statistical method is more optimal to implement in the problem, Kolmogorov–Smirnov, Anderson–Darling, or perhaps a combination (hybrid) of these two approaches?

In this proof-of-the-concept paper, we will explore this new method by predicting direct targets for bacterial pleiotropic



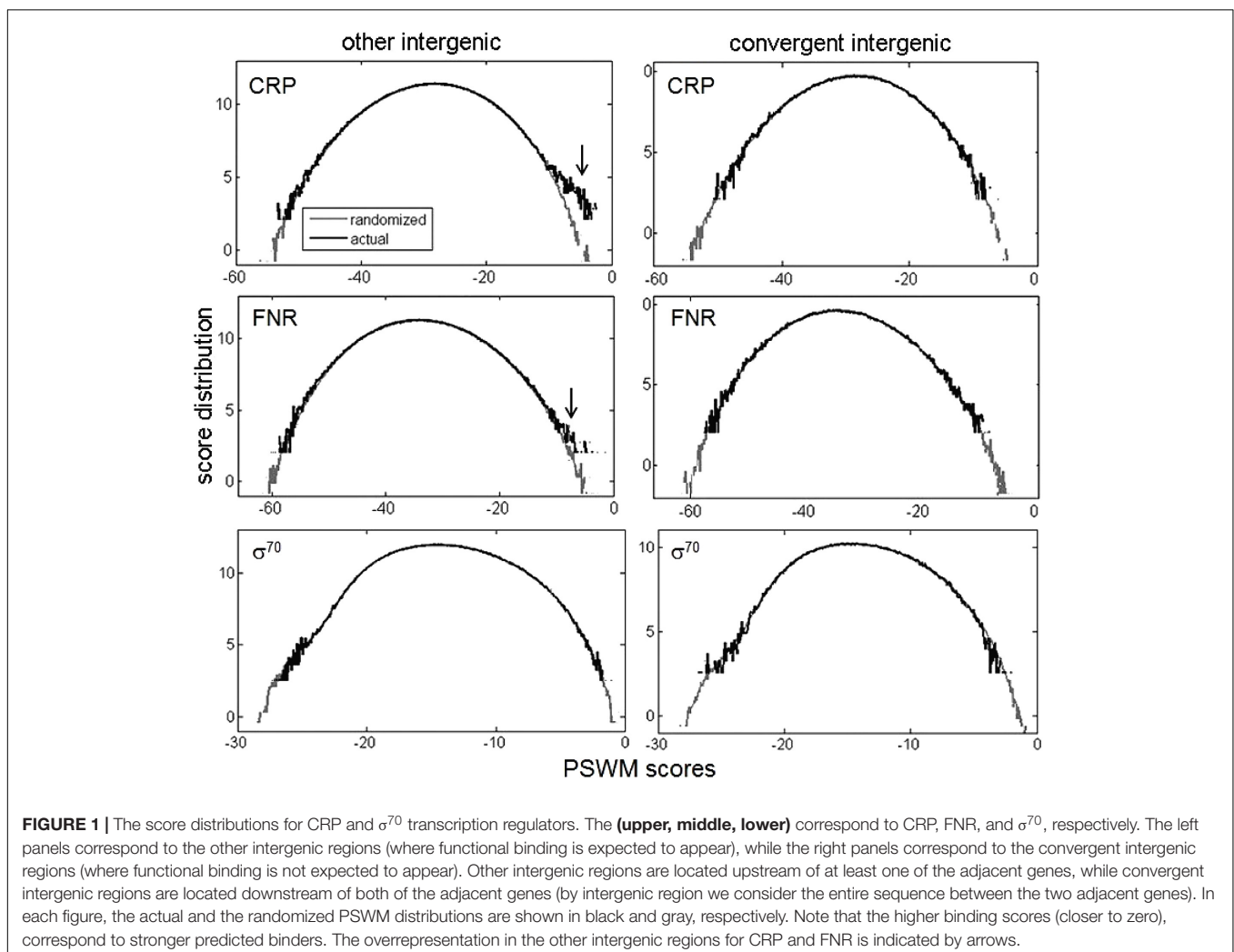
regulators ( $\sigma^{70}$ , CRP, FNR), which present a classical (currently unresolved) bioinformatics problem characterized by low prediction specificity. On the other hand, accurately predicting transcription targets of bacterial regulators is crucial for understanding bacterial gene expression regulation. Considering bacterial regulators also allows a more straightforward interpretation of the obtained results, as complicating issues such as chromatin state/accessibility (Forties et al., 2011; Chen and Bundschuh, 2014; Chereji and Morozov, 2014) that are present in eukaryotes are largely absent here.

## RESULTS AND DISCUSSION

### Overrepresentation of PSWM Scoring Distributions

We start by exploring the basic concept behind our hypothesis that the distribution of PSWM scores is overrepresented in the regions where binding of transcription regulators is expected, and that the overrepresentation is absent in the regions where they do not bind. This concept is illustrated by the upper panel

of **Figure 1**, where binding of a pleiotropic *Escherichia coli* transcription factor CRP (also known as CAP) to the convergent intergenic regions, and to the rest of the intergenic regions (here called the “other intergenic regions”), is assessed. Note that the convergent intergenic regions are located downstream of both of the adjacent genes, while the other intergenic regions are located upstream of at least one of the adjacent genes. Therefore, there should be no CRP binding sites in the convergent intergenic regions, while CRP binding sites should be located in a subset of the other intergenic regions, which are upstream of its regulatory targets. Accordingly, in the upper left panel of **Figure 1**, we observe a significant overrepresentation of CRP PSWM scores in the other intergenic regions, while such overrepresentation is absent in the convergent intergenic regions. Note that, in **Figure 1**, the background distribution corresponds to randomized intergenic regions, with the sequences randomized so as to preserve trinucleotide frequencies. We obtain similar results (the middle panels) for another *E. coli* pleiotropic transcription factor (FNR), i.e., we also observe an overrepresentation in the other intergenic regions (though now smaller compared to CRP), and an



absence of overrepresentation in the convergent intergenic regions.

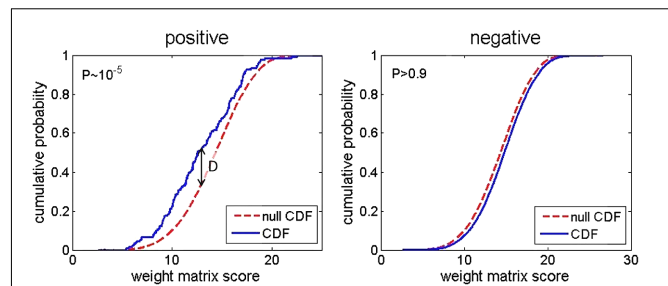
On the other hand, a more complex case is presented in the lower panels of **Figure 1**. Here, binding of the *E. coli*  $\sigma^{70}$  factors to the other intergenic (the left panel) and the convergent intergenic (the right panel) regions is assessed. Note that bacterial  $\sigma$  factors ensure transcription initiation (i.e., provide signal for transcription start sites), and different  $\sigma$  factors are related with transcription exhibited under different conditions in a bacterium (Paget and Helmann, 2003; Feklistov et al., 2014). In particular,  $\sigma^{70}$  is the housekeeping  $\sigma$  factor in *E. coli*, which is associated with transcribing a large number of bacterial genes under normal conditions (therefore having a large regulon). We observe an absence of overrepresentation in both the other and the convergent intergenic regions, in fact a small underrepresentation in the high scoring tail for the convergent intergenic regions can be observed. The absence of the overrepresentation is likely a consequence of significant negative selection on  $\sigma^{70}$  non-sites, as a subset of the other intergenic regions (from which transcription of the downstream genes is directed) has to be enriched with  $\sigma^{70}$  binding sites.

$\sigma^{70}$  binding, in which no global overrepresentation is observed, evidently corresponds to a more complex case of the regulatory target recognition. Consequently, in the results below, we will first concentrate on  $\sigma^{70}$ , to demonstrate utility of the method even in a more complicated scenario. In addition, prediction of  $\sigma$  factor binding sites, and their corresponding direct targets (i.e., genes that they transcribe) is a classical (unresolved) bioinformatics problem that is considered notoriously hard (Stormo, 2000; Towsey et al., 2008; Purto et al., 2014), but one that is crucial for understanding bacterial transcription. Predictions of  $\sigma^{70}$  targets are moreover important since RNA-seq experiments (which can map transcription start-sites) are still rare in bacteria, and a number of transcription start sites are active under non-standard conditions, which likely differ from those used in the experiments (Feklistov et al., 2014). Therefore, an additional motivation is to investigate whether our approach can lead to reasonable predictions for such a difficult problem. We will then come back to analyzing two other *E. coli* pleiotropic regulators (CRP and FNR), which display the more standard/expected binding score distributions.

## Kolmogorov–Smirnov Based Approach

The main idea behind the new approach is to observe an overrepresentation of PSWM score distribution for the entire upstream genomic region of interest, with respect to a chosen background (null) distribution. We then need to provide a measure of the difference between the two scoring distributions (corresponding to the upstream genomic regions, and the background distribution), as well as a measure of statistical significance for this difference. Assessing this difference can be directly implemented through Kolmogorov–Smirnov (KS) test, which is illustrated in **Figure 2**.

In the left panel, an example of an upstream intergenic region, which is clearly enriched by  $\sigma^{70}$  binding sites, is shown. The solid curve corresponds to the cumulative distribution function (CDF), corresponding to PSWM scores of this intergenic region.

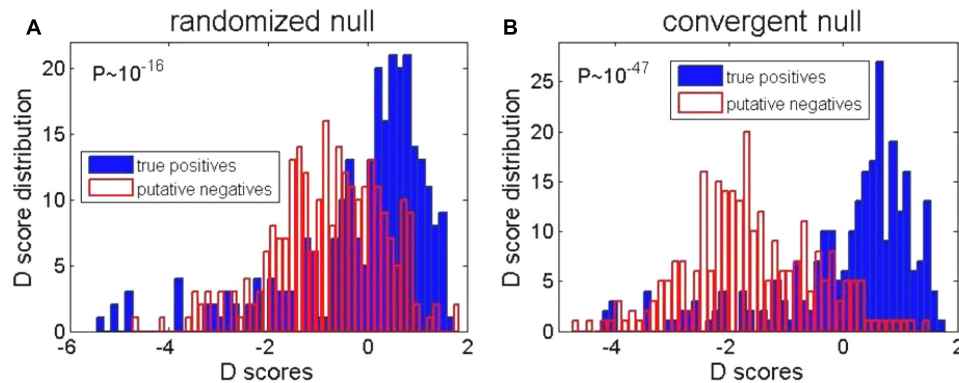


**FIGURE 2 |** Cumulative distribution functions corresponding to PSWM scores for the upstream genomic regions (the solid blue curve), and the background distribution (the dashed red curve) are shown in the figure. In the (left), an intergenic region that is strongly enriched by  $\sigma^{70}$  binding sites, and that would be reported as a direct target in the search, is shown. In the (right), an extreme case of the upstream intergenic region, which is depleted of  $\sigma^{70}$  binding sites, is shown. *P*-values from KS are indicated in both panels.

Note that the usual KS measure of the difference between the two distributions (which we here denote as *D* score) is indicated in the figure. With respect to the *D* score, note that we here use the one-sided KS test, i.e., we impose the condition that CDF of the upstream genomic regions has to be above the background distribution CDF, which is the condition that corresponds to overrepresentation – i.e., the case of significant underrepresentation being reported as a hit is excluded. KS test also directly provides the *P*-value corresponding to this *D* score, which in turn allows assessing statistical significance of the potential target. On the other hand, the right panel presents an example where the upstream intergenic region is depleted of  $\sigma^{70}$  binding sites. In this case, CDF of PSWM scores corresponding to this depleted intergenic region is actually below the background distribution, so that the gene downstream of this intergenic region is clearly not reported as a direct target of  $\sigma^{70}$  (*D* score is very close to zero in this case). Note that CDF of the upstream intergenic region does not have to be below the background CDF (as happens in the extreme case shown in the right panel), to be excluded as a hit. That is, all hits with small *D* values, which are statistically non-significant, are not reported as putative targets.

## Enrichment of *D* Scores

To implement the KS based method, the choice of the background (null) distribution becomes important. This is actually already indicated in **Figure 1**, where we have seen that, due to the negative selection, the distribution corresponding to the randomized regions may not overlap with the distribution in the regions where no binding happens. We here test two choices of the background distributions: (A) the distribution corresponding to the randomized regions, where the intergenic regions are randomized, and their corresponding PSWM scoring distribution is used as the background, (B) genomic regions where functional binding is not expected, for which we use the convergent intergenic regions, as explained in **Figure 1**. Note that these two choices correspond, respectively, to the left (the randomized regions) and the right (the convergent intergenic regions) panel shown in **Figure 3**.



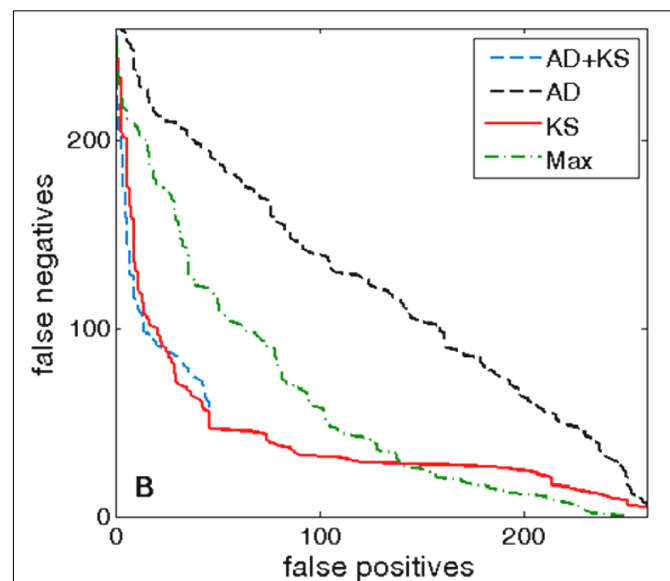
**FIGURE 3 |** Enrichment of  $D$  scores for two background distributions. The (A,B) represent, respectively, the background distributions that correspond to the randomized and the convergent intergenic regions. The open red and the solid blue histograms, respectively, correspond to the  $D$  score distributions for the positives (the upstream intergenic regions with experimentally detected binding sites) and the putative negatives (the genomic regions deep inside *E. coli* ORF, where functional  $\sigma^{70}$  binding does not appear). The difference between the blue and the red distributions is assessed by the  $P$ -value, indicated in each panel.

For each of these two choices of the background distributions, the  $D$  score distribution is calculated in the following two cases: (i) the red histogram, which corresponds to positives (i.e., the upstream intergenic regions, which are experimentally known to contain  $\sigma^{70}$  binding sites); (ii) the blue histogram: which corresponds to putative negatives, i.e., the genomic regions where  $\sigma^{70}$  binding should *not* appear. Specifically, we here use genomic sequences deep inside ORF (coding sequences), where we expect no initiation of transcription (i.e., no functional  $\sigma^{70}$  binding). We here mark such regions as putative negatives.

We see a significant enrichment of  $D$  scores in the true positive vs. putative negative regions, for both choices of the background distributions (i.e., for both **Figures 3A,B**). However, we see that the enrichment is clearly much higher when the background distribution corresponds to the convergent intergenic regions, as clearly indicated by the  $P$ -values in the **Figures 3A,B**. The most likely reason is that the randomized regions do not capture (possibly significant) negative selection that acts on  $\sigma^{70}$  binding sites. That is, the functional binding, which one needs to detect, comes on the ‘top’ of possibly a large number of non-sites that are ‘deleted’ by the negative selection. Consequently, in the further analysis, we will use the background distribution which corresponds to the convergent intergenic regions.

## ROC Curves and Comparison with Anderson–Darling Test

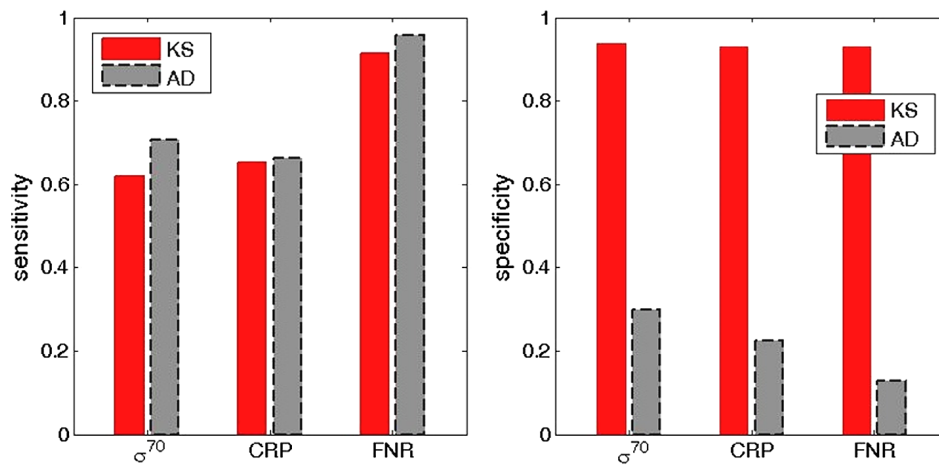
Our next goal is to compare the accuracy of KS-based approach, with the standard method for identifying putative targets in bacteria. This method (which we further call “Max”) involves scanning the upstream genomic regions by PSWM, and classifying as putative targets those regions that contain individual binding sites with PSWM scores above certain threshold. To this end, we use the same positives and putative negatives as introduced in the previous subsection, and the null distribution that corresponds to the convergent intergenic regions. In addition, as an alternative to KS test, the AD test can also be used to detect overrepresentation of the binding



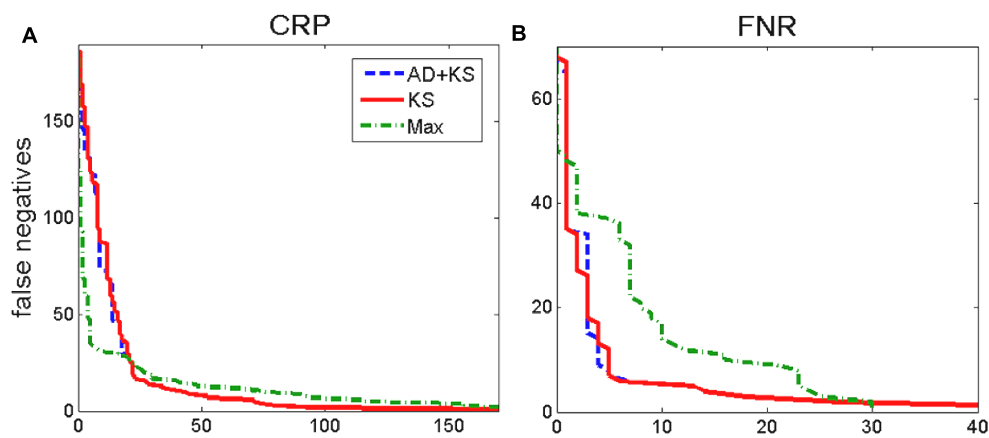
**FIGURE 4 |** ROC curves,  $\sigma^{70}$  binding. The ROC curves for  $\sigma^{70}$  direct targets, which correspond to different tests, are marked in the legend. The positive and the putative negative sets are the same as in **Figure 3**, and the false negatives and the false positives are estimated with respect to these two sets, where the convergent background distribution is used. In the legend, “Max” corresponds to the standard method for direct target identification (see the first paragraph of this subsection).

scores with respect to the null distribution. Consequently, we also address how accurately the two tests (AD and KS) can predict direct targets of  $\sigma^{70}$ . The corresponding prediction accuracies are assessed by ROC curves shown in **Figure 4**.

Importantly, we see that KS based approach (the solid red curve) shows a substantially better performance compared to the standard method (the dot-dashed green curve). In particular, note that for the fixed number of false positives there are up to three times fewer false negatives. Such a reduction in the



**FIGURE 5 |** Comparison of sensitivity and specificity for KS and AD methods. The (left, right) correspond, respectively, to the sensitivity and specificity estimates, obtained for the usual  $P = 0.05$  confidence level. The red and the gray bars correspond to KS and AD methods respectively. The sensitivity and the specificity estimates are shown for  $\sigma^{70}$  (the left bars), CRP (the central bars), and FNR (the right bars). The sensitivity and specificity are calculated as, respectively,  $TP/P$ , and  $TN/N$ , where TP are true positives, TN true negatives, while P and N are the number of positives and negatives, respectively.



**FIGURE 6 |** ROC curves, CRP and FNR. The (A) and the (B) correspond, respectively, to the ROC curves for CRP and FNR. The true positive and the putative negatives are as in Figure 3, and the convergent intergenic regions are used for the background distribution. Different tests which correspond to the shown ROC curves are marked in the figure legend.

number of false positives is expected, i.e., in accordance with the hypothesis we presented above, since individual sites with high PSWM scores can easily appear by random. On the other hand, their appearance is automatically taken into account through the background distribution, i.e., a potential target will be classified as a hit only if the binding scores in the entire searched region are enriched (overrepresented) with respect to the background distribution.

Furthermore, we see that KS (the solid red curve) leads to a higher detection accuracy compared to AD (the dashed black curve). Moreover, KS test is also much ( $\sim 400$  times) faster in predicting the direct targets. Consequently, in this application, KS test is both faster and more accurate than AD. Note that this runs opposite to the common paradigm, according to which AD is slower, but more accurate compared to KS (Stephens, 1974).

To investigate the reason behind the (unexpected) significantly higher accuracy obtained with KS method, in Figure 5 we compare the sensitivity (the left panel) and the specificity (the right panel) for KS and AD methods. The comparison corresponds to the standard classification threshold ( $P < 0.05$ ) for both methods, and is provided for  $\sigma^{70}$  (analyzed in Figure 4) and for CRP and FNR transcription factors (analyzed in Figures 6, 7 below). We see that the sensitivity is high, and about the same, for both methods (with AD displaying even slightly larger sensitivity). On the other hand, in the right panel of Figure 5, it can be seen that the specificity is much smaller for AD method, which then leads to its lower accuracy compared to KS method. To interpret this result, one should note that we necessarily work with an approximation of the true null distribution, e.g., the negative selection on non-sites in the

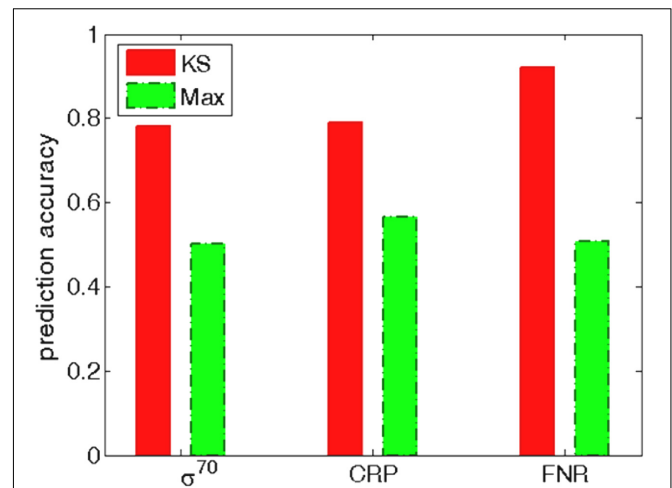


convergent intergenic regions is likely not the same as in the other intergenic regions, in which the target classification is exhibited. Consequently, the main general advantage of AD method, which is its large sensitivity, becomes a weakness in this application, as small differences with respect to the null distribution (that may also arise from its approximate nature), are (conveniently) not captured by KS, but are classified as statistically highly significant by AD test, leading to low AD specificity (a large number of false positives).

Next, there comes a question if a combination of AD and KS tests can provide an improved accuracy compared to either of the two tests alone. With that respect, we made an algorithm corresponding to a hybrid where the KS test is implemented first to filter-out those upstream regions with clearly insignificant  $P$ -values, i.e., regions where the actual distribution is clearly too close to the null distribution. Afterward, AD is applied to those regions with distributions that are more different to the null distribution, for which AD performs better. From **Figure 4**, we see that such AD-KS hybrid (the dashed blue curve) indeed shows a substantially better accuracy compared to AD test alone, and has a similar accuracy to KS test alone. This result is consistent with the discussion above, i.e., when AD test alone is used, a number of the upstream regions that are eliminated by KS test, are falsely classified as targets by AD (since, due to small specificity, AD proclaims even small difference between the distributions as being significant). On the other hand, when in AD-KS hybrid AD is applied only to those distributions that are more different with respect to the null distribution (therefore bypassing its main problem of low specificity), the accuracy becomes similar to KS test. As an outlook, note that AD-KS hybrid might be further improved by optimizing the threshold for KS selection. We will further concentrate only on KS and AD-KS tests, as they have a much better accuracy to AD test alone.

We next come back to assessing KS approach for two more standard binding score distributions (see **Figure 1**), exhibited by CRP and FNR transcription regulators. We here construct the positive and the putative negative sets in the same way as for  $\sigma^{70}$ , i.e., the positives correspond to the intergenic regions where the transcription regulator binding is experimentally shown, while the putative negatives correspond to the sequences deep inside the coding regions, where functional binding is not expected. The corresponding ROC curves are shown in **Figure 6**.

For FNR (**Figure 6B**), we obtain similar results as for  $\sigma^{70}$ , i.e., KS (and KS-AD hybrid) lead to a significantly better ROC curve performance compared to the standard method (e.g., for a fixed false negative number, there is a several times smaller number of false positives for KS). On the other hand, we see that for CRP the two curves (KS and the standard method) have apparently similar performances, i.e., while the standard method shows better performance at low false positive numbers, it is outperformed by KS at higher false positives. The similar performance of KS in the case of CRP is not surprising, i.e., is likely a consequence of the fact that functional binding dominates over non-sites in this case, as implied by the large PSWM score overrepresentation exhibited in such case (see the upper left panel in **Figure 1**). Consequently, the results in **Figure 5** are in line with



**FIGURE 7 |** Prediction accuracy for the standard classification thresholds. The threshold in KS search is based on the estimated statistical significance (the usual  $P = 0.05$  threshold is taken). The threshold in the PSWM search corresponds to the standard choice where most (98%) of the experimentally determined binding examples would be recovered in the search. The prediction accuracy for these two thresholds is shown for  $\sigma^{70}$  (the left bars), CRP (the central bars), and FNR (the right bars). The prediction accuracy is calculated as  $(TP + TN)/(TP + FP + FN + TN)$ , where TP (true positives), TN (true negatives), FP (false positives) and FN (false negatives) are calculated for the two methods at the corresponding threshold choices.

our main hypothesis that the main utility of KS approach is in accurate classification of non-sites.

## Statistical Significance and the Classification Threshold

Independently from the ROC performance, KS has a significant advantage of straightforwardly assigning statistical significance to each predicted target, which is normally not available for standard PSWM search (see Introduction). We here explore the utility of such robust statistical significance estimate with the example of assigning a classification threshold. With the KS approach a natural threshold choice is provided by the  $P$ -value, typically set to  $P = 0.05$ . As such a natural choice is normally not available for standard PSWM search, the threshold is usually set so that almost all ( $\sim 98\%$ ) of the experimentally determined binding sites from which PSWM is constructed are recovered in the search. In **Figure 7**, we explore the search accuracy associated with the two choices of the binding threshold, i.e.,  $P = 0.05$  for KS method and the standard threshold (see above) for PSWM search.

We see that the threshold based on KS significance estimate leads to much higher prediction accuracy for  $\sigma^{70}$  and FNR, which is expected based on the significantly better ROC performance of KS in these two cases (**Figures 4, 6B**). Moreover, in **Figure 7** we also see notably higher prediction accuracy in the case of CRP, where a similar ROC performance was observed for KS and standard PSWM search (**Figure 6A**). Consequently, the notably higher search accuracy for KS in the case of CRP observed in **Figure 7** is based on the more optimal choice of the classification

threshold. This underlines the advantage of the threshold choice based on the robust statistical significance measure.

## CONCLUSION AND OUTLOOK

We here proposed a new computational approach to direct regulatory target prediction. The approach is based on assessing the significance of the difference between PSWM scoring distributions, which correspond to the upstream genomic regions, and the background distribution. As a consequence, P-value is assigned to the entire upstream region, instead of searching for individual high-scoring binding sites. We implemented this approach through classical Kolmogorov–Smirnov and Anderson–Darling tests, as well as through a hybrid of these two approaches. Surprisingly, and contrary to the current paradigm, we have seen that the approach based on Kolmogorov–Smirnov test leads to a higher search accuracy compared to Anderson–Darling based approach, while also being (as expected) computationally less demanding. While the hybrid approach has a substantially higher accuracy compared to Anderson–Darling test, it does not outperform the simpler Kolmogorov–Smirnov test. We interpreted this result by Anderson–Darling test classifying small differences with respect to the background distribution as true binding targets, leading to low specificity of the approach.

We furthermore showed that the Kolmogorov–Smirnov based approach leads to a substantially higher accuracy compared to the standard approach, reducing the number of false positives for several times. Moreover, a clear advantage of Kolmogorov–Smirnov approach is that it straightforwardly assigns statistical significance to any tested upstream intergenic region. We demonstrated this advantage on the example of the classification threshold, where we have seen that the robust significance estimate provided by Kolmogorov–Smirnov leads to a much more optimal threshold choice. We find that genomic regions, where functional binding is not expected, provide better background compared to randomized genomic regions. We here, i.e., for analysis of prokaryotic transcription regulation, used convergent intergenic regions for background distribution. In eukaryotes the choice of background distribution would be more complicated and remains to be investigated, where one possibility would be to take genomic sequences far from coding regions (where there may not be much TFBS).

To prove this new concept in the direct regulatory target prediction, we tested it in the case of pleiotropic bacterial regulators. This allowed a more straightforward interpretation of the obtained results, while testing the method on some of the classical problems otherwise characterized by low prediction specificity. As an outlook, the method proposed here is of a general significance, and it will be in the future also implemented in the more complicated case of direct target prediction for eukaryotic transcriptional regulators. Moreover, while the model was here applied in the context of PSWM, more complex models which take into account interdependences of nucleotides in TFBS were also developed (Eggeling et al., 2015; Kulakovskiy et al., 2016; Nettling et al., 2017). While these methods lead to a

better performance in some cases, more often (simpler) PSWMs perform better, which is likely due to overfitting, i.e., due to a limited number of TFBS from which the model is trained (Benos et al., 2002; Nguyen and Androulakis, 2009). Therefore, despite the limitations of PSWMs, they are still the leading approach in TFBS search (Nguyen and Androulakis, 2009; Fazius et al., 2011). In any case, the new approach proposed here does not depend on the scoring method (i.e., if a classical PSWM, or a higher order model, is used), since the approach is based on comparing the distributions of the scores (i.e., is not limited by how the actual scores are calculated). Consequently, the KS approach proposed here might present a general method of choice for efficiently and accurately predicting target loci of transcription regulators.

## MATERIALS AND METHODS

### Defining the Upstream Genomic Regions

The *E. coli* intergenic sequences are divided in two groups, where binding of transcription regulators is expected (other intergenic regions) and not expected (convergent intergenic regions). The other intergenic regions, and the convergent intergenic regions, include, respectively, those that are located upstream of at least one adjacent gene, and downstream of both of the adjacent genes.

For the positive set in  $\sigma^{70}$  case, in KS, AD and KS-AD hybrid searches, we take those *E. coli* intergenic sequences that contain  $\sigma^{70}$  binding sites with experimental evidence from RegulonDB database (Gama-Castro et al., 2011), which results in the total of 263 upstream genomic regions. Similarly, for the positive set in CRP and FNR case, we take these intergenic sequences with experimental evidence of the regulator binding from RegulonDB database. For the putative negative set, we use the same number of sequences, with the same length, as those in the positive set, but now sampled from ORF (coding sequences), where we exclude 50 bps at both 5' and 3' ends; this is done to exclude the flanking sequences, in which  $\sigma^{70}$  binding sites are sometimes located.

For obtaining the randomized distribution, an ensemble of randomized sequences was constructed, by sampling all trinucleotide probabilities in the intergenic regions. The randomized sequences were searched, and the corresponding randomized scoring distributions are obtained, in the same manner as for the upstream genomic regions, which is further described below.

### PSWM Scoring Distributions

CRP and FNR PSWM were constructed from the binding sites assembled in DPInteract database (Robison et al., 1998), through the standard information-theory based procedure (Stormo, 2000).  $\sigma^{70}$  PSWM were constructed starting from recent *de novo* alignment (Djordjevic, 2011), where the promoter elements were systematically aligned starting directly from the experimentally determined TSS. Briefly, the alignment includes: –10 and –35 elements, spacer weights corresponding to variable spacer length (between 15 and 19 bps), conserved sequences upstream of –10 element. PSWMs for CRP, FNR and  $\sigma^{70}$  search are provided in Supplementary Table 1. Scores were assigned to each DNA segment in the upstream genomic and the randomized sequences

using these PSWM, from which the corresponding scoring distributions were generated.

## KS, AD, and KS-AD Hybrid Based Searches

For KS based search, one-sided Kolmogorov–Smirnov test was used. For each tested upstream intergenic region PSWM score distribution was generated as described above, and compared with an appropriate background distribution whose CDF was constructed. This comparison results in *P*-value and *D* score for each tested upstream genomic segment. The threshold on *D* scores was then moved in order to change the number of false positives and false negatives, and construct the ROC curves.

For AD based search, the MATLAB based routine ‘adtest’ was used, where PSWM score distributions corresponding to upstream genomic region, and the background distribution were compared. For each tested upstream genomic region, *P*-value and AD test statistics (‘adstat’) was sampled. The ROC curves were constructed based on ‘adstat’ scores.

For KS-AD hybrid search, KS and AD tests were implemented as described above, with KS test used first to exclude the upstream intergenic regions with low difference between the two distributions. A liberal *P*-value threshold of 0.5 was used in this exclusion, so that only the upstream genomic regions with very low significance are eliminated by KS test. The rest of the upstream regions are then subjected to AD test, which is used to calculate *P*-value and AD test statistics. The codes for KS, AD and KS-AD hybrid approaches are available upon request.

## REFERENCES

- Benos, P. V., Bulyk, M. L., and Stormo, G. D. (2002). Additivity in protein–DNA interactions: how good an approximation is it? *Nucleic Acids Res.* 30, 4442–4451. doi: 10.1093/nar/gkf578
- Bulyk, M. L. (2004). Computational prediction of transcription-factor binding site locations. *Genome Biol.* 5:201. doi: 10.1186/gb-2003-5-1-201
- Bulyk, M. L. (2006). DNA microarray technologies for measuring protein–DNA interactions. *Curr. Opin. Biotechnol.* 17, 422–430. doi: 10.1016/j.copbio.2006.06.015
- Chen, C., and Bundschuh, R. (2014). Quantitative models for accelerated protein dissociation from nucleosomal DNA. *Nucleic Acids Res.* 42, 9753–9760. doi: 10.1093/nar/gku719
- Chereji, R. V., and Morozov, A. V. (2014). Functional roles of nucleosome stability and dynamics. *Brief. Funct. Genomics* 14, 50–60. doi: 10.1093/bfpg/elu038
- de Jong, A., Pietersma, H., Cordes, M., Kuipers, O. P., and Kok, J. (2012). PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* 13:299. doi: 10.1186/1471-2164-13-299
- Djordjevic, M. (2011). Redefining *Escherichia coli*  $\sigma^{70}$  promoter elements: –15 motif as a complement of the –10 motif. *J. Bacteriol.* 193, 6305–6314. doi: 10.1128/JB.05947-11
- Djordjevic, M. (2013). Efficient transcription initiation in bacteria: an interplay of protein–DNA interaction parameters. *Integr. Biol.* 5, 796–806. doi: 10.1039/c3ib20221f
- Djordjevic, M., and Sengupta, A. M. (2006). Quantitative modeling and data analysis of SELEX experiments. *Phys. Biol.* 3, 13–28. doi: 10.1088/1478-3975/3/1/002

## AUTHOR CONTRIBUTIONS

All authors have given approval to the final version of the manuscript. MarD conceived the work, with the help of MagD and EZ. MarD and MagD implemented the method and performed the analysis. All the authors interpreted the results. MarD wrote the paper, with the help of MagD and EZ.

## FUNDING

This work is supported by the Swiss National Science Foundation under SCOPES project number IZ73Z0\_152297, by Marie Curie International Reintegration Grant within the 7th European Community Framework Program (PIRG08-GA-2010-276996), and by the Ministry of Education and Science of the Republic of Serbia under Project number ON173052.

## ACKNOWLEDGMENTS

We thank Djordje Markovic and Marija Basic for help with **Figure 1**, and to Christopher Rands for critically reading the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02314/full#supplementary-material>

- Djordjevic, M., Sengupta, A. M., and Shraiman, B. I. (2003). A biophysical approach to transcription factor binding site discovery. *Genome Res.* 13, 2381–2390. doi: 10.1101/gr.1271603
- Eggeling, R., Roos, T., Myllymäki, P., and Grosse, I. (2015). Inferring intra-motif dependencies of DNA binding sites from ChIP-seq data. *BMC Bioinformatics* 16:375. doi: 10.1186/s12859-015-0797-4
- Favorov, A. V., Gelfand, M. S., Gerasimova, A. V., Ravcheev, D. A., Mironov, A. A., and Makeev, V. J. (2005). A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length. *Bioinformatics* 21, 2240–2245. doi: 10.1093/bioinformatics/bti336
- Fazius, E., Shelest, V., and Shelest, E. (2011). SiTaR: a novel tool for transcription factor binding site prediction. *Bioinformatics* 27, 2806–2811. doi: 10.1093/bioinformatics/btr492
- Feklistov, A., Sharon, B. D., Darst, S. A., and Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* 68, 357–376. doi: 10.1146/annurev-micro-092412-155737
- Forties, R. A., North, J. A., Javadi, S., Tabbaa, O. P., Fishel, R., Poirier, M. G., et al. (2011). A quantitative model of nucleosome dynamics. *Nucleic Acids Res.* 39, 8306–8313. doi: 10.1093/nar/gkr422
- Gama-Castro, S., Jiménez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spinola, M. I., Contreras-Moreira, B., et al. (2008). RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation. *Nucleic Acids Res.* 36(Suppl. 1), D120–D124. doi: 10.1093/nar/gkm994
- Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñoz-Rascado, L., Solano-Lira, H., et al. (2011). RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic

- sensory response units (Sensor Units). *Nucleic Acids Res.* 39(Suppl. 1), D98–D105. doi: 10.1093/nar/gkq1110
- Green, P. J., Kay, S. A., Lam, E., and Chua, N.-H. (1989). “In vitro DNA footprinting,” in *Plant Molecular Biology Manual*, eds S. B. Gelvin, R. A. Schilperoort, and B. P. S. Verma (Dordrecht: Springer), 415–436. doi: 10.1007/978-94-009-0951-9\_21
- Hannenhalli, S. (2008). Eukaryotic transcription factor binding sites—modeling and integrative search methods. *Bioinformatics* 24, 1325–1331. doi: 10.1093/bioinformatics/btn198
- Hertz, G. Z., and Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* 15, 563–577. doi: 10.1093/bioinformatics/15.7.563
- Jagannathan, V., Roulet, E., Delorenzi, M., and Bucher, P. (2006). HTPSELEX—a database of high-throughput SELEX libraries for transcription factor binding sites. *Nucleic Acids Res.* 34, D90–D94. doi: 10.1093/nar/gkj049
- Jayaram, N., Usvyat, D., and Martin, A. C. (2016). Evaluating tools for transcription factor binding site prediction. *BMC Bioinformatics* doi: 10.1186/s12859-016-1298-9 [Epub ahead of print].
- Kim, T. H., and Ren, B. (2006). Genome-wide analysis of protein-DNA interactions. *Annu. Rev. Genomics Hum. Genet.* 7, 81–102. doi: 10.1146/annurev.genom.7.080505.115634
- Korostevy, Y. D., Zharov, I. A., Mironov, A. A., Rakhmaininova, A. B., and Gelfand, M. S. (2016). Identification of position-specific correlations between DNA-binding domains and their binding sites. Application MerR family transcription factors. *PLOS ONE* 11:e0162681. doi: 10.1371/journal.pone.0162681
- Kulakovskiy, I. V., Vorontsov, I. E., Yevshin, I. S., Soboleva, A. V., Kasianov, A. S., Ashoor, H., et al. (2016). HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.* 44, D116–D125. doi: 10.1093/nar/gkv1249
- Levitsky, V. G., Kulakovskiy, I. V., Ershov, N. I., Oshchepkov, D. Y., Makeev, V. J., Hodgman, T., et al. (2014). Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data. *BMC Genomics* 15:80. doi: 10.1186/1471-2164-15-80
- Locke, G., and Morozov, A. V. (2015). A biophysical approach to predicting protein–DNA binding energetics. *Genetics* 200, 1349–1361. doi: 10.1534/genetics.115.178384
- Mathelier, A., Fornes, O., Arenillas, D. J., Chen, C.-Y., Denay, G., Lee, J., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–D115. doi: 10.1093/nar/gkv1176
- Nettling, M., Treutler, H., Cerquides, J., and Grosse, I. (2017). Combining phylogenetic footprinting with motif models incorporating intra-motif dependencies. *BMC Bioinformatics* 18:141. doi: 10.1186/s12859-017-1495-1
- Newburger, D. E., and Bulky, M. L. (2009). UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. *Nucleic Acids Res.* 37(Suppl. 1), D77–D82. doi: 10.1093/nar/gkn660
- Nguyen, T. T., and Androulakis, I. P. (2009). Recent advances in the computational discovery of transcription factor binding sites. *Algorithms* 2, 582–605. doi: 10.3390/a2010582
- Ozoline, O. N., and Deev, A. A. (2006). Predicting antisense RNAs in the genomes of *Escherichia coli* and *Salmonella typhimurium* using promoter-search algorithm PlatProm. *J. Bioinform. Comput. Biol.* 4, 443–454. doi: 10.1142/S0219720006001916
- Paget, M. S. B., and Helmann, J. D. (2003). The  $\sigma^{70}$  family of sigma factors. *Gen. Biol.* 4, 203–208. doi: 10.1186/gb-2003-4-1-203
- Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–680. doi: 10.1038/nrg2641
- Purtov, Y. A., Glazunova, O. A., Antipov, S. S., Pokusaeva, V. O., Fesenko, E. E., Preobrazhenskaya, E. V., et al. (2014). Promoter islands as a platform for interaction with nucleoid proteins and transcription factors. *J. Bioinform. Comput. Biol.* 12:1441006. doi: 10.1142/S0219720014410066
- Robison, K., McGuire, A., and Church, G. (1998). A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.* 284, 241–254. doi: 10.1006/jmbi.1998.2160
- Roulet, E., Busso, S., Camargo, A., Simpson, A., Mermod, N., and Bucher, P. (2002). High-throughput SELEX–SAGE method for quantitative modeling of transcription-factor binding sites. *Nat. Biotechnol.* 20, 831–835. doi: 10.1038/nbt718
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* 69, 730–737. doi: 10.1080/01621459.1974.10480196
- Stormo, G. D. (2000). DNA binding sites: representation and discovery. *Bioinformatics* 16, 16–23. doi: 10.1093/bioinformatics/16.1.16
- Stormo, G. D., and Fields, D. S. (1998). Specificity, free energy and information content in protein–DNA interactions. *Trends Biochem. Sci.* 23, 109–113. doi: 10.1016/S0968-0004(98)01187-6
- Stormo, G. D., and Zhao, Y. (2010). Determining the specificity of protein–DNA interactions. *Nat. Rev. Genet.* 11, 751–760. doi: 10.1038/nrg2845
- Towsey, M., Hogan, J., Mathews, S., and Timms, P. (2008). The in silico prediction of promoters in bacterial genomes. *Genome Inform.* 19, 178–189.
- Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510. doi: 10.1126/science.2200121
- Vilar, J. M. (2010). Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. *Biophys. J.* 99, 2408–2413. doi: 10.1016/j.bpj.2010.08.006
- Vilar, J. M., and Saiz, L. (2013). Systems biophysics of gene expression. *Biophys. J.* 104, 2574–2585. doi: 10.1016/j.bpj.2013.04.032
- Wade, J. T., Struhl, K., Busby, S. J., and Grainger, D. C. (2007). Genomic analysis of protein–DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol. Microbiol.* 65, 21–26. doi: 10.1111/j.1365-2958.2007.05781.x
- Wingender, E. (2008). The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.* 9, 326–332. doi: 10.1093/bib/bbn016

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Djordjevic, Djordjevic and Zdobnov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Successful Establishment of Plasmids R1 and pMV158 in a New Host Requires the Relief of the Transcriptional Repression of Their Essential *rep* Genes

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Matxalen Llosa,  
University of Cantabria, Spain  
Elisabeth Grohmann,  
Beuth University of Applied Sciences,  
Germany  
Antonio Juárez,  
University of Barcelona, Spain

### \*Correspondence:

Gloria del Solar  
gdelsolar@cib.csic.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 August 2017

**Accepted:** 16 November 2017

**Published:** 01 December 2017

### Citation:

Ruiz-Masó JA, Luengo LM,  
Moreno-Córdoba I, Díaz-Orejías R and  
del Solar G (2017) Successful  
Establishment of Plasmids R1 and  
pMV158 in a New Host Requires the  
Relief of the Transcriptional  
Repression of Their Essential *rep*  
Genes. *Front. Microbiol.* 8:2367.  
doi: 10.3389/fmicb.2017.02367

José Á. Ruiz-Masó, Luis M. Luengo, Inmaculada Moreno-Córdoba, Ramón Díaz-Orejías and Gloria del Solar\*

Molecular Microbiology and Infection Biology Department, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

Although differing in size, encoded traits, host range, and replication mechanism, both narrow-host-range theta-type conjugative enterobacterial plasmid R1 and promiscuous rolling-circle-type mobilizable streptococcal plasmid pMV158 encode a transcriptional repressor protein, namely CopB in R1 and CopG in pMV158, involved in replication control. The gene encoding CopB or CopG is cotranscribed with a downstream gene that encodes the replication initiator Rep protein of the corresponding plasmid. However, whereas CopG is an auto-repressor that inhibits transcription of the entire *copG-repB* operon, CopB is expressed constitutively and represses a second, downstream promoter that directs transcription of *repA*. As a consequence of the distinct regulatory pathways implied by CopB and CopG, these repressor proteins play a different role in control of plasmid replication during the steady state: while CopB has an auxiliary role by keeping repressed the regulated promoter whenever the plasmid copy number is above a low threshold, CopG plays a primary role by acting coordinately with RNAIL. Here, we have studied the role of the regulatory circuit mediated by these transcriptional repressors during the establishment of these two plasmids in a new host cell, and found that excess Cop repressor molecules in the recipient cell result in a severe decrease in the frequency and/or the velocity of appearance of transformant colonies for the cognate plasmid but not for unrelated plasmids. Using the pMV158 replicon as a model system, together with highly sensitive real-time qPCR and inverse PCR methods, we have also analyzed the effect of CopG on the kinetics of repopulation of the plasmid in *Streptococcus pneumoniae*. We show that, whereas in the absence of CopG pMV158 repopulation occurs mainly during the first 45 min following plasmid transfer, the presence of the transcriptional repressor in the recipient cell severely impairs the replicon repopulation

and makes the plasmid replicate at approximately the same rate as the chromosome at any time after transformation, which results in maximal plasmid loss rate in the absence of selection. Overall, these findings indicate that unrepressed activity of the Cop-regulated promoter is crucial for the successful colonization of the recipient bacterial cells by the plasmid.

**Keywords:** plasmid repopulation, establishment phase replication, R1 replicon, pMV158 replicon, Cop transcriptional repressors, plasmid replication rate

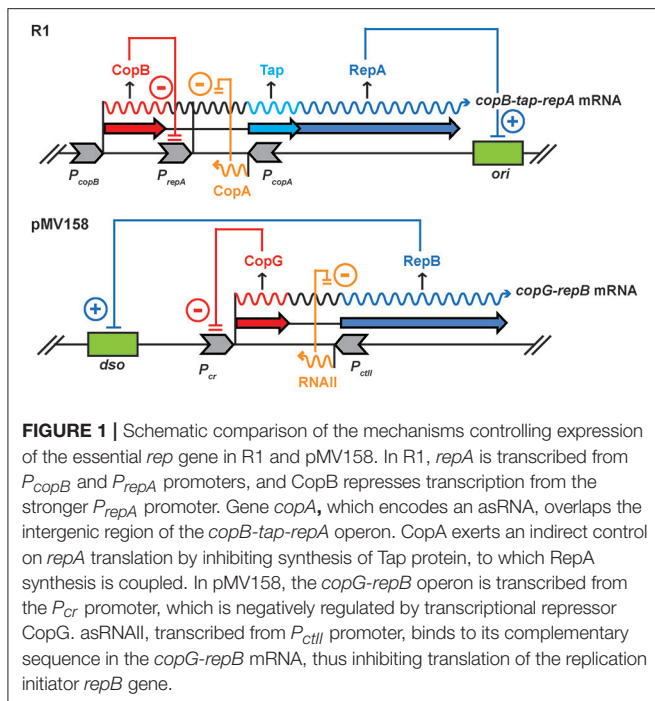
## INTRODUCTION

Plasmids specify replication control systems that enable them to maintain a characteristic steady-state concentration (copy number) in their host cell. These regulatory systems are *trans-acting* and can sense and correct stochastic up and down fluctuations of the plasmid copy number in individual cells. By adjusting the replication initiation rate in response to changes in the intracellular plasmid concentration, control systems manage to keep the steady-state condition, where every plasmid copy replicates, on average, once per cell generation (Nordström, 1990). Steady-state control of plasmid replication has been analyzed at a deep level and a variety of different regulatory circuits have been mechanistically characterized (del Solar and Espinosa, 2000; Das et al., 2005; Nordström, 2006). Antisense RNA (asRNA)-mediated control of plasmid replication is widely spread in theta- and rolling circle-replicating plasmids from Gram-positive and Gram-negative bacteria (del Solar and Espinosa, 2000; Brantl, 2014). The small regulatory RNAs involved in control of plasmid replication and copy number are *bona fide* (i.e., *cis*-encoded) asRNAs, as they are encoded on the DNA strand opposite to an RNA essential for replication initiation, namely a pre-primer or an mRNA for the replication initiator protein (Rep). These asRNAs base-pair to their target (sense) RNA to inhibit its function and/or the completion of its synthesis through a variety of mechanisms, including inhibition of primer maturation, transcription attenuation, prevention of formation of a translation activator RNA pseudoknot, and inhibition of translation of either *rep* or a leader-peptide reading frame to which *rep* is translationally coupled (del Solar and Espinosa, 2000; Brantl, 2014). Most frequently, asRNAs controlling plasmid replication are metabolically unstable, *trans-acting* inhibitory elements, whose synthesis is directed by unregulated and strong promoters. These features enable them to sense and correct rapidly up and down fluctuations of the plasmid copy number in individual cells (del Solar and Espinosa, 2000; Wagner et al., 2002; Brantl, 2014).

Although the asRNA is the sole replication control element in some plasmids (pT181 family, IncB/IncI $\alpha$  family, ColE2), it is accompanied, in others, by a regulatory protein that acts either as a transcriptional repressor (Cop proteins in R1, Inc18, and pMV158 families) or as an RNA-binding protein (Rom/Rop protein in ColE1-like plasmids) (del Solar and Espinosa, 2000; Brantl, 2014). Rom/Rop and CopB of ColE1- and R1-like replicons, respectively, have been largely considered as mere auxiliary elements because of their rather secondary role in the steady-state plasmid replication control, when the activity

of these proteins is almost saturating (Nordström et al., 1984; Rosenfeld and Grover, 1993; Atlung et al., 1999; Summers, 2009). In contrast, efficient replication control of plasmids of the Inc18 and pMV158 families requires the coordinated participation of the asRNA and of the transcriptional repressor Cop protein, both elements playing a primary regulatory role (del Solar and Espinosa, 2000; Brantl, 2014).

Plasmid R1, originally isolated from *Salmonella enterica* serovar Paratyphi, is a low-copy-number, multiresistance, conjugative plasmid of the IncFII incompatibility group. It has a narrow host-range restricted to the *Enterobacteriaceae* family. The R1 elements and circuits involved in steady-state plasmid replication control and maintenance have been studied in great detail (Nordström et al., 1984; Olsson et al., 2004; Nordström, 2006). In addition to the origin of replication (*oriR1*), the R1 basic replicon includes the *repA* gene for the replication initiator protein and the two replication control genes, *copB* and *copA*, encoding, respectively, the transcriptional repressor CopB protein and the CopA asRNA (**Figure 1**). The RepA protein is rate limiting for initiation of replication. The essential *repA* gene is transcribed from two promoters, namely *P<sub>copB</sub>* and *P<sub>repA</sub>*. The upstream *P<sub>copB</sub>* promoter directs constitutive transcription of *copB*, *tap* (the leader peptide reading frame) and *repA*, which is translationally coupled to *tap*. Transcription from the downstream CopB-regulated *P<sub>repA</sub>* promoter gives rise to the shorter bicistronic *tap-repA* mRNA (**Figure 1**). When unrepressed, the *P<sub>repA</sub>* promoter is about twice as strong as *P<sub>copB</sub>*, although under normal conditions during the steady-state plasmid replication, *P<sub>repA</sub>* is almost totally (90%) switched off by CopB-mediated repression (Olsson et al., 2004). Since CopB acts as a tetramer (Riise and Molin, 1986), the activity of *P<sub>repA</sub>* is likely to be strongly dependent on plasmid concentration. It has been shown that the presence of extra copies of *copB* *in trans*, which further reduces the already low activity of *P<sub>repA</sub>*, increases 7-fold the rate of loss of a Par<sup>+</sup> derivative of the R1 basic replicon (Olsson et al., 2004). The steady-state activity of *P<sub>repA</sub>* is thought to stabilize the plasmid inheritance both by speeding up R1 replication in cells with very few plasmid copies (thus decreasing the frequency of these cells) and by slightly increasing the average plasmid concentration. CopB also plays a main role in the coupling between the *kis-kid* auxiliary maintenance system and the basic replicon of the plasmid (López-Villarejo et al., 2015). The switch of this coupling is the antitoxin Kis, whose levels decrease in cells with lower-than-average plasmid copy number. Decrease of Kis concentration activates the Kid toxin, which is an RNase with two efficient target sites in the intergenic region of the *copB-repA* mRNA (Pimentel et al., 2005). Cleavage



at these sites reduces the CopB levels, which leads to activation of  $P_{repA}$  and subsequent increase in plasmid replication efficiency. Hence, the *kis-kid* system coupled to the CopB-mediated loop functions as a safety device when the plasmid copy number is very low.

Promiscuous plasmid pMV158 was originally isolated from a clinical strain of *Streptococcus agalactiae* (Burdett, 1980) and subsequently transferred to a large number of bacterial genera and species. The detailed analysis of its replicon has made pMV158 the prototype of a vast family of rolling circle-replicating plasmids (Ruiz-Masó et al., 2015; Boer et al., 2016). The pMV158 basic replicon includes a compact region containing the double-strand origin (*dso*) (Ruiz-Masó et al., 2007) as well as the genes that encode the replication initiator protein (RepB) (Ruiz-Masó et al., 2004; Boer et al., 2009) and the two replication control elements (transcriptional repressor CopG and asRNA RNAII) (del Solar et al., 1995; Gomis-Rüth et al., 1998; Hernández-Arriaga et al., 2009; López-Aguilar et al., 2015) (Figure 1). Like the *rep* gene product of R1, RepB of plasmid pMV158 is the rate-limiting factor for the replication initiation process. Unlike R1, the streptococcal plasmid expresses the essential *rep* gene from a single promoter ( $P_{cr}$ ), which directs cotranscription of the *copG-repB* operon and is subjected to CopG-mediated regulation (del Solar et al., 1990) (Figure 1). In the unrepressed state,  $P_{cr}$  seems to be even stronger than promoter  $P_{cIII}$  that directs synthesis of countertranscript RNAII, a situation that contrasts with the plasmid replication control systems based exclusively on asRNA, where the transcription rate of the essential RNA is constant but rather low compared with that of the asRNA (del Solar and Espinosa, 2000). Unsuccessful repression of promoter  $P_{cr}$  in pMV158 derivatives encoding a defective CopG repressor

leads to a 5-fold increase in the plasmid copy number (del Solar et al., 1990, 1995). On the other hand, the presence *in trans* of high dosages of the autoregulated *copG* gene has been shown to decrease by ~35% the steady-state copy number of the pMV158 replicon in *S. pneumoniae* (del Solar et al., 1995).

Despite the quite deep current knowledge about the involvement of the Cop transcriptional repressors in the control of the steady-state plasmid replication, very little has been reported so far on their role in the establishment phase replication. The Cop regulatory elements have been proposed to play an important role during plasmid establishment in a new bacterium based on the fact that the Cop-regulated promoter, when unrepressed, determines high transcription rates of the essential *rep* gene (del Solar et al., 1990, 1995; del Solar and Espinosa, 2000; Olsson et al., 2004; Brantl, 2014). Yet, only non-published results have been invoked in a few articles to suggest the involvement of the Cop regulatory loops of pMV158 and R1 in the establishment phase replication of these plasmids (Nordström and Nordström, 1985; del Solar and Espinosa, 2000; Olsson et al., 2004).

In this work, we have analyzed the effect of the Cop proteins of R1 and pMV158 on the establishment of these plasmids. To this end, we have electrotransferred a mini-R1 derivative (pKN1562) to *Escherichia coli* and *Salmonella Typhimurium* (*S. Typhimurium* hereafter) cells that either contain or lack a compatible recombinant plasmid encoding CopB. Similarly, we have transferred pLS1 (a *mob*<sup>-</sup> derivative of pMV158, Lacks et al., 1986) to naturally-competent pneumococcal cells and to electrocompetent cells of *Staphylococcus aureus* either containing or lacking a compatible recombinant plasmid encoding CopG. We show that, irrespective of the system employed, the presence of these proteins in the recipient cell selectively impairs the establishment of the cognate plasmid, resulting in a decrease in the frequency of total or early transformant colonies. By using the pneumococcal host as a model system, we also show that repopulation of the pMV158 replicon is almost abolished when autoregulated *copG* is supplied *in trans* at a high gene dosage.

## MATERIALS AND METHODS

### Bacterial Strains and Plasmids

Bacterial strains employed, and their uses for this work, are summarized in Table 1. Plasmid constructions used throughout this study, as well as their relevant features, are listed in Table 2. *S. pneumoniae* 708 was the host for pLS1, pLS1*cop7*, pC194, pCGA3, pCGA3n, pCGA30, and pAMβ1 plasmids. Pneumococcal cells were grown at 37°C in AGCH medium (Lacks et al., 1986) supplemented with 0.3% sucrose and 0.2% yeast extract. *S. aureus* RN4220 was the host for pLS1, pT181*cop608*, pC194, pCGA3, pCGA3n, and pCGA30 plasmids. Staphylococcal cells were grown at 37°C in brain heart infusion medium (BHI, Difco). *E. coli* C600 and *S. Typhimurium* SL1344 were the hosts for pUC18-*copB*, pKN1562, and pACYC184 plasmids; cells were grown at 30°C in Lysogeny broth (LB) medium.

**TABLE 1** | Strains used in this study.

Bacteria		
Bacteria	Characteristics	Source
<i>Escherichia coli</i> C600	K12 derivate	Appleyard, 1954
<i>Salmonella</i> Typhimurium SL1344	Wild-type strain	Hoiseth and Stocker, 1981 NC_016810.1
<i>Staphylococcus aureus</i> RN4220	Restriction-deficient mutant of strain 8325-4	Kreiswirth et al., 1983
<i>Streptococcus pneumoniae</i> 708	<i>end-1</i> <i>exo-1</i> <i>trt-1</i> <i>hex-4</i> <i>malM594</i>	Lacks and Greenberg, 1977

## Plasmidic and Genomic DNA Preparations

Plasmidic DNA (pDNA) content from pneumococcal transformants was analyzed by preparing total DNA crude extracts as described (del Solar et al., 1987). These DNA preparations were also used to estimate the relative plasmid copy number from the ratio between the intensities of the plasmid and chromosome DNA bands quantified for the plasmid of interest relative to a plasmid control whose copy number has been precisely determined, after correcting for the difference in size of both plasmids (del Solar et al., 1995). Plasmids pLS1, pLS1*cop7*, pC194, pCGA3, pCGA3n, and pCGA30 were isolated from *S. pneumoniae* 708 and purified by two consecutive CsCl/ethidium bromide density gradient centrifugations, as described (Lacks et al., 1986). Plasmid pAMβ1 was isolated from *S. pneumoniae* 708 and purified by alkaline lysis as described (Stassi et al., 1981). Plasmid pT181*cop608*, isolated from *S. aureus* RN4220, and plasmids pKN1562, pUC18-*copB*, and pACYC184, isolated from *E. coli* C600, were purified using a Jetstar Plasmid Midiprep Kit (Genomed). pDNA content from staphylococcal transformants was analyzed by the same alkaline lysis method used for *S. pneumoniae*. In both midipreps and alkaline lysis procedures lysostaphin (50 µg/ml) was added to the cell resuspension buffer in order to facilitate staphylococcal cells lysis.

Genomic DNA (gDNA) used as template for real-time quantitative PCR (qPCR) and inverse PCR (iPCR) was isolated from pneumococcal cultures in exponential growth phase, which was determined by measurement of optical density at 650 nm. The DNA was extracted from cells of *S. pneumoniae* 708 with different plasmid content by using the Wizard® Genomic DNA Purification Kit (Promega) optimized for *S. pneumoniae*. Cells resuspended in 50 mM EDTA were incubated with 0.04% of deoxycholate and 0.1 mg/ml of Proteinase K for 10 min at 37°C. Next, and before proceeding with the lysis step, the cellular suspension was quickly frozen on a mixture of dry ice and ethanol and stored at −80°C. With this method, aliquots taken at different time intervals were processed simultaneously from the lysis step. Moreover, 0.05 µg/ml of glycogen (molecular biology grade) was added in the isopropanol precipitation step to facilitate gDNA recovery from diluted samples. Contrarily to the samples where no glycogen was added, the gDNA yield of the samples treated with glycogen was found to be nearly proportional to the total amount of lysed cells. Concentration of the gDNA was determined with a Qubit fluorometer by using the Qubit HS dsDNA Assay Kit (Molecular Probes).

Purified gDNA was digested with EcoRI, a restriction enzyme that linearizes the pLS1 DNA but leaves intact the plasmidic and chromosomal amplicons (i.e., the DNA segments to be amplified in the qPCR assays). This method has been developed to obtain accurate qPCR-based copy number results for plasmids (Providenti et al., 2006).

## Calculation of the Experimental Plasmid Loss Rate

The experimental loss rate ( $L_{ex}$ ) of pLS1 and pLS1*cop7* in newly transformed pneumococcal cells was calculated from the equation (Gerdes et al., 1985):

$$T/T_0 = (1 - L_{ex})^n, \quad (1)$$

where  $T_0$  and  $T$  are, respectively, the fractions of transformants *ab initio* and after  $n$  generations. This equation can be converted into a linear function by taking logarithms,

$$\log(T/T_0) = \log(1 - L_{ex})n, \quad (2)$$

where  $\log(1 - L_{ex})$  is the slope of the linear regression fit in the plot of the experimental values of  $\log(T/T_0)$  against the number of cell generations ( $n$ ).

## Transformation of Bacterial Species with Plasmid DNA

Transformation of *E. coli* C600 and *S. Typhimurium* SL1344 cells was performed by electroporation essentially as described (Dower et al., 1988). Competent cultures of *E. coli* C600 and *S. Typhimurium* SL1344 and those of the same strains harboring pUC18-*copB* as the resident plasmid were transformed with 0.2 µg of plasmid DNA of pKN1562 or pACYC184. Transformants were grown on LB-agar plates with antibiotic selection according to the resistance carried by the plasmids: 50 µg/ml of kanamycin (Km) for pKN1562, 50 µg/ml of ampicillin (Amp) for pUC18-*copB*, or 20 µg/ml of chloramphenicol (Cm) for pACYC184. Competent cells of *S. aureus* RN4220 were prepared and transformed by electroporation following the procedure depicted in Augustin and Götz (1990). Competent staphylococcal cultures (50 µl) harboring pC194, pCGA3, pCGA3n, or pCGA30 as the resident plasmid were transformed with 0.5 µg of DNA of the donor plasmid (pLS1 or pT181). After allowing for phenotypic expression (60 min), all cultures were treated for 30 min with 0.5 µg/ml of tetracycline (Tc), a sub-inhibitory concentration of the antibiotic that allows induction of the pT181 *tet* gene. Cells transformed with pLS1 or pT181 were selected on BHI-agar plates containing 5 µg/ml of Tc, for selection of the entering plasmid, and 3 µg/ml of Cm, for resident plasmid selection.

Competent cells of *S. pneumoniae* 708 were prepared and transformed as described (López et al., 1982). Three independent lots of naturally competent cells were prepared from each of the four different strains harboring pC194, pCGA3, pCGA3n, or pCGA30 as the resident plasmid. Since the development of pneumococcal competence is influenced by many factors (Attaiech et al., 2015), including the exact composition of the



**TABLE 2 |** Plasmids used in this study.

Plasmids				
Plasmids	Characteristics	Antibiotic marker	Copy number	Source
pACYC184	Plasmid vector containing the minimal p15A replicon	Cm <sup>R</sup> Tc <sup>R</sup>	≈15 <sup>a</sup>	Chang and Cohen, 1978
pAMβ1	Conjugative plasmid from <i>Enterococcus faecalis</i> DS5	Em <sup>R</sup>	≈7 <sup>c</sup>	Leblanc and Lee, 1984
pCGA3	<i>copG</i> cloned in the high copy number vector pC194	Cm <sup>R</sup>	≈200 <sup>e</sup>	del Solar et al., 1995
pCGA3n	<i>copG</i> cloned in the medium copy number vector pC194n	Cm <sup>R</sup>	≈25 <sup>e</sup>	del Solar et al., 1995
pCGA30	pCGA3 derivative with a 48-bp deletion that includes the—10 sequence of the <i>P<sub>Cr</sub></i> promoter, and the <i>SD</i> sequence and the two first codons of <i>copG</i> .	Cm <sup>R</sup>	≈200 <sup>e</sup>	del Solar et al., 1993
pC194	Rolling circle replicating plasmid from <i>S. aureus</i>	Cm <sup>R</sup>	≈200 <sup>e</sup>	del Solar et al., 1995
pKN1562	Wild-type mini-R1 plasmid carrying the wt <i>kis-kid</i> system	Km <sup>R</sup>	1–2	Molin et al., 1979
pLS1	Non-mobilizable pMV158 derivative plasmid ( $\Delta mobM$ , $\Delta ssuU$ )	Tc <sup>R</sup>	46.0 ± 5.0 <sup>c,d</sup>	Stassi et al., 1981
pLS1 <i>cop7</i>	pLS1 copy-up mutant	Tc <sup>R</sup>	182.1 ± 15.5 <sup>c,d</sup>	del Solar et al., 1990
pT181- <i>cop608</i>	pT181-based high copy number mutant plasmid	Tc <sup>R</sup>	800–1,000 <sup>b</sup>	Rasooly et al., 1994
pUC18- <i>copB</i>	<i>copB</i> cloned in the high copy number vector pUC18	Amp <sup>R</sup>	≈500 <sup>a</sup>	López-Villarejo et al., 2015

Plasmid copy number (PCN) determined in *E. coli* C600<sup>a</sup>, *S. aureus*<sup>b</sup>, and *S. pneumoniae*<sup>c</sup>, <sup>d</sup>PCN determined by RT-qPCR in this work (± standard error), <sup>e</sup>Approximate PCN value estimated by comparative quantitation of total DNA extracted from pneumococcal strains containing the target plasmid or the control pLS1 plasmid; PCN of the latter has been determined by RT-qPCR.

semi-defined AGCH culture medium, the 12 competent cultures were each tested for their level of competence by transformation with chromosomal DNA from a strain able to grow in maltose. Although the transformation efficiency for maltose utilization varied significantly from lot to lot of competent cells of the same strain, the level of competence for chromosomal transformation of the competent cells prepared in parallel was, consistently, 1.5 to 2-, 2 to 3-, and 4 to 5-fold higher for the strains containing pCGA30, pCGA3n, and pCGA3, respectively, than for the strain harboring pC194. The reason for the apparent increase in the natural competence of *S. pneumoniae* when the DNA of gene *copG* is present has not been investigated yet. Cultures (1 ml) of competent pneumococcal cells were transformed with 0.25 µg of DNA of the donor plasmid (pLS1 or pAMβ1). After allowing for phenotypic expression (70 min), cultures were induced with 0.5 µg/ml of Cm for 20 min. Transformants were selected using agar plates containing 1 µg/ml of Tc for pLS1 or 1 µg/ml of erythromycin for pAMβ1. In these plates, selection for the resident plasmid (3 µg/ml of Cm) was maintained. Since pneumococci grow best when protected from air, the basal AGCH-agar layer containing the cells was overlaid with AGCH-agar medium.

## Repopulation Kinetics Assays

For repopulation kinetics experiments, cultures of competent pneumococcal cells harboring pCGA3 and pCGA30 were subjected to a modified version of the transformation procedure described in López et al. (1982) that yielded 10-fold higher competence levels. Competent cultures (OD<sub>650</sub> = 0.3) were diluted 1/20 in 10 ml of AGCH medium supplemented with 0.3% of sucrose, 0.001% of CaCl<sub>2</sub> and a sub-inhibitory concentration of Cm (0.5 µg/ml), in order to keep the induced expression of the *cat* gene. The cells were cultured at 37°C to an OD<sub>650</sub> of 0.3, and the cultures were cooled to 30°C for 15 min. Then, the cells were transformed with 2 µg of DNA of the pMV158

derivative (pLS1 or pLS1*cop7*) by incubation for 30 more min at the same temperature. To stop the transformation process, pancreatic DNase I was added to a final concentration of 2 µg/ml, and the incubation at 30°C was prolonged for 20 more min. Next, the cultures were diluted 1/10 in pre-warmed (37°C) AGCH medium supplemented with 0.3% sucrose, 0.2% yeast extract and 0.5 µg/ml of Cm, and incubated at 37°C up to 150 min. Immediately after dilution, and at the indicated time intervals, 10 and 0.1-ml aliquots of the cultures were withdrawn and used, respectively, to extract the gDNA and to determine the number of total viable cells (c.f.u./ml) and the fraction of transformants. Transformants were selected in three-layered AGCH-medium agar plates containing 3 µg/ml of Cm and 1 µg/ml of Tc. Cells were deposited in the basal layer that was overlaid with a second layer of AGCH-agar medium. The plates were then incubated at 37°C for 2 h before antibiotics for selection were included in the third layer and spread across the rest of the plate by diffusion.

No-transformation control experiments were also performed to ensure that plasmid DNA amplified by real-time qPCR arose from the transformed cells and was not contaminant DNA that escaped from the DNase I digestion. For this purpose, we followed the same transformation procedure as described above but adding simultaneously 2 µg of pLS1*cop7* DNA and DNase I (2 µg/ml), in order to avoid transformation. Aliquots of 10 ml of the no-transformed cultures were taken immediately after dilution and after 30 min of incubation at 37°C. These cell aliquots were processed as described to obtain gDNA. In all cases, before proceeding with the gDNA isolation protocol, cells were washed with 10 ml of 1X PBS (phosphate-buffered saline).

The total number of cell generations (*n*) was calculated according to the following equation:

$$n = \log(V/V_0)/\log 2, \quad (3)$$

where  $V$  is the number of viable cells (c.f.u.) at any of the times analyzed and  $V_0$  is the initial value of viable cells.

A similar expression was used to calculate the number of gDNA duplications ( $D_C$ ) at a given time interval ( $t_i - t_j$ ):

$$D_{C_{i-j}} = \log(gDNA_j / gDNA_i) / \log 2, \quad (4)$$

where  $gDNA_j$  and  $gDNA_i$  are the amounts (ng) of gDNA obtained (after precipitation in the presence of glycogen, see above) at the times  $t_j$  and  $t_i$ , respectively.

### Determination of the Copy Number of a Specific pMV158 Amplicon Relative to a Chromosomal Amplicon in Transformed Pneumococcal Cells by qPCR

Two primer sets specific to the PcrA helicase single-copy reference gene (*pcrA*) of *S. pneumoniae* R6 (Hoskins et al., 2001) and to the tetracycline resistance TetL protein gene (*tetL*) of pMV158 were designed. Oligonucleotide primers sets (Table 3) were designed with Primer3 v0.4.0 (Koressaar and Remm, 2007; Untergasser et al., 2012) based on the pLS1 sequence (NC\_010096.1) and on the *S. pneumoniae* R6 (NC\_003098.1) *pcrA* sequence. Criteria used during primer design were that primers had predicted  $T_m$  of  $\sim 59^\circ\text{C}$  and that they generated amplicons  $\sim 140$  bp in length.

qPCRs were conducted in a total volume of 20  $\mu\text{l}$  using a LightCycler<sup>®</sup> 96 real-time detection system (Roche) and the FastStart Essential DNA Green Master (Roche), as per manufacturer's recommendations. Decimally diluted EcoRI-digested gDNA preparations (14, 1.4, 0.14 ng per reaction) were analyzed using 0.5  $\mu\text{M}$  (final concentration) of the specific forward and reverse primers of either primer-pair used (Table 3). To prepare the reactions and minimize pipetting errors 2  $\mu\text{l}$  of template DNA were added to individual qPCRs. Thermal cycling conditions were as follows: initial denaturation at  $95^\circ\text{C}$  for 5 min, followed by 40 cycles of  $95^\circ\text{C}$  for 10 s (denaturation),  $59^\circ\text{C}$  for 30 s (primer annealing), and  $72^\circ\text{C}$  for 20 s (elongation). A melting curve analysis of the PCR products, with a temperature gradient of  $0.1^\circ\text{C}/\text{s}$  from 59 to  $95^\circ\text{C}$ , was performed to confirm the purity and specificity of the PCR products. Two independent qPCR trials were conducted for each template source. In each trial, triplicate samples of the three different amounts of template were analyzed. Control samples without template DNA were also analyzed.

Relative copy number (CN) of the pMV158 amplicon was calculated using equation:

$$CN = (1 + E_{pcrA})^{Ct_{pcrA}} / (1 + E_{tetL})^{Ct_{tetL}}, \quad (5)$$

where  $E_{pcrA}$  and  $E_{tetL}$  are, respectively, the PCR amplification efficiencies of the chromosomal and plasmid amplicons, and  $Ct_{pcrA}$  and  $Ct_{tetL}$  are the mean threshold cycle values obtained for the corresponding amplicons. A CN value was calculated for each of the three template concentrations analyzed, and the mean and standard deviation of the six values (two independent trials with three different template concentrations each) were determined.

$E$  values of target ( $E_{tetL}$ ) and reference ( $E_{pcrA}$ ) sequences were empirically calculated for each qPCR trial. For that purpose, mean  $Ct$  values were plotted against the logarithm of the amount of total DNA template in the assay. From the slope of the curve generated by linear regression of the plotted points, the PCR amplification efficiency was determined according to the equation:

$$E = 10^{-1/\text{slope}} - 1, \quad (6)$$

Although the  $E$  values for both amplicons were higher than 0.9, we have chosen Equation (5) to calculate the relative copy number of the plasmid amplicon as it allows taking into account the slight differences between  $E_{target}$  and  $E_{reference}$  that we have observed.

### Determination of the Relative Amount of Circular Plasmid DNA in Transformed Pneumococcal Cells by iPCR

The plasmidic DNA (pDNA) present in the gDNA isolated from the transformed pneumococcal cells was used as template to perform an inverse PCR protocol with a primer set of divergent oligonucleotides. iPCR was performed using the Phusion High Fidelity (HF) (Thermo Scientific) DNA polymerase. Amplification reactions (20  $\mu\text{l}$ ) contained 0.7 ng of gDNA and 0.5  $\mu\text{M}$  of the specific forward and reverse primers (Table 3). Thermal cycling conditions comprised 25 cycles ( $98^\circ\text{C}$  for 10 s,  $59.5^\circ\text{C}$  for 30 s, and  $72^\circ\text{C}$  for 1 min and 25 s) plus a final extension step of 10 min at  $72^\circ\text{C}$ . The amplification reaction yielded a linear dsDNA fragment corresponding to almost the entire pLS1cop7 plasmid. The products of iPCR were analyzed on 0.8% agarose gels, stained with GelRed (Biotium), and quantified with the aid of a Gel Doc (BIO-RAD) system. At least three gels with DNA products obtained in each of three independent iPCR assays were analyzed. *In vitro* DNA amplification in these iPCR assays was based on equation:

$$P = P_0(1 + E)^C, \quad (7)$$

where  $P$  and  $P_0$  are, respectively, the amount of amplified linear pDNA product and the initial amount of template pDNA in the gDNA used for the amplification reaction;  $E$  is the amplification efficiency, and  $C$  is the number of cycles. Irrespective of the gDNA concentration used, the ratio between  $P$  and  $P_0$  is kept constant for a given  $C$  provided there is no exhaustion of the primers and dNTPs required for DNA synthesis (and hence  $E$  is kept constant). We then confirmed that the employed iPCR conditions fulfilled this requirement for gDNA concentrations ranging from half to twice that used for the analysis of the kinetics of plasmid repopulation.

On the other hand, we have defined the relative plasmid amplification occurring in the transformants during the time interval  $t_i - t_j$  as the ratio between the relative numbers of plasmid molecules (with respect to the total gDNA) at times  $t_j$  and  $t_i$ , which can be estimated from the intensity of the bands corresponding to the linear pDNA products obtained in the iPCR assays using the same amount of gDNA extracted at different

**TABLE 3** | Oligonucleotides used in this study.

Target	Accession No.	Sequence (5′-3′) <sup>a</sup>	Length (nt)	Primer position	Product size (bp)
Sequence of primers used for real-time qPCR					
<i>pcrA</i>	NC_003098.1	F: GAGTTGGTTGAGTCCGTCCT	20	980570–980589	144
		R: TGTCACATCCGTGGTGCAT	20	980713–980694	
<i>tetL</i>	NC_010096.1	F: TGCGAGTACAACTGGGTGA	20	1879–1898	146
		R: ACCCAATTACCGACCCGAAA	20	2024–2005	
Sequence of primers used for iPCR					
<i>del-5</i>	NC_010096.1	F: GTTTGAGGCTCGTCAAATC	19	514–532	4403
<i>dso2</i>	NC_010096.1	R: CAGCTCTAAGGCTAAAGGCG	20	508–489	

<sup>a</sup>F and R indicate forward and reverse primers, respectively.

times after transformation. Considering that, in the time interval  $t_i$  to  $t_j$ , the plasmid replication rate (i.e., the ratio of pDNA duplications,  $D_p$ , to gDNA duplications,  $D_c$ ) has a value of  $R$ , the relative plasmid amplification in this interval is given by the following equation:

$$Pl_{t_j}/Pl_{t_i} = 2^{D_p}/2^{D_c} = 2^{(R D_c - D_c)} = 2^{D_c(R-1)}, \quad (8)$$

where  $Pl_{t_j}$  and  $Pl_{t_i}$  are the relative intensities of the amplified pDNA products at times  $t_j$  and  $t_i$ , respectively. Equation (8) can be converted into a linear function by taking logarithms:

$$\log(Pl_{t_j}/Pl_{t_i}) = D_c(R-1) \log 2, \quad (9)$$

therefore, the  $R$  value in the time interval  $t_i$  to  $t_j$  was calculated from the equation:

$$R = (\log(Pl_{t_j}/Pl_{t_i})/D_c \log 2) + 1 \quad (10)$$

## Statistical Analysis

ANOVA was run to determine whether experimental Q ratios differed among groups of staphylococcal strains ( $p$ -values < 0.05 were considered significant).

## RESULTS

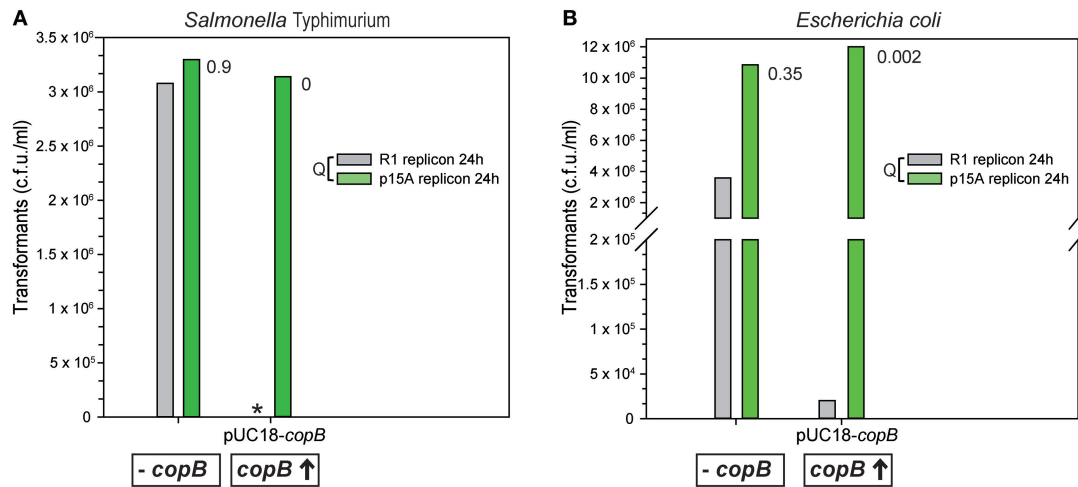
### The Presence of the Cop Repressor Protein of R1 or pMV158 in the Recipient Cell Decreases the Frequency and/or the Velocity of Appearance of Colonies Transformed with the Cognate Plasmid

To know whether an initially unrepressed transcription of the essential *rep* gene is required for successful establishment of plasmids R1 and pMV158 in a new cell, we compared the efficiencies with which recipient cells that contain or lack the Cop transcriptional repressor of either plasmid were transformed in parallel with plasmids harboring the cognate replicon or an unrelated replicon. It is worth noting that the incoming plasmids carry their own *cop* genes, and hence these assays aim to analyze the importance of the recipient cells having the Cop protein already synthesized upon plasmid entrance.

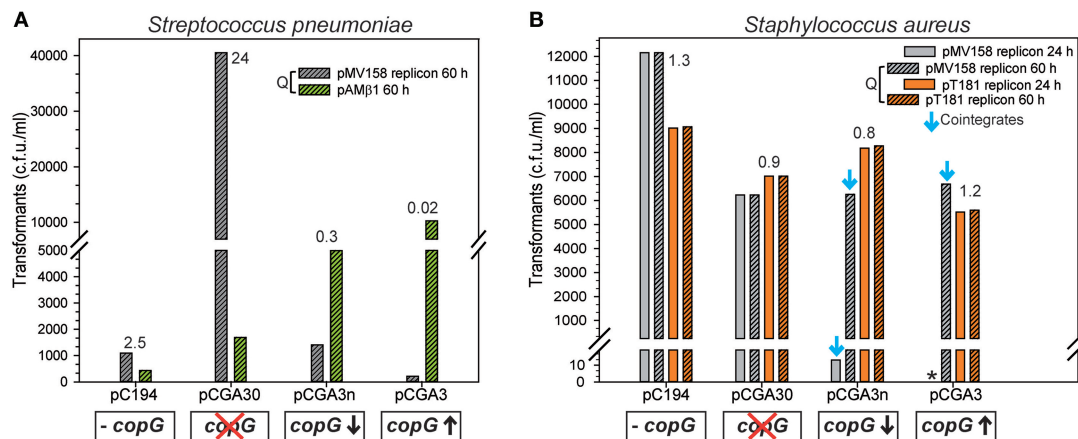
The existence of a specific effect of R1 CopB on the establishment of the cognate plasmid was tested by transforming either plasmid-free or pUC18-*copB*-carrying *E. coli* C600 and *S. Typhimurium* SL1344 (WT) cells with DNAs of plasmids pKN1562 (a mini-R1 derivative) or pACYC184 (harboring the R1-unrelated p15A replicon), both of which are compatible with the pUC18 replicon. Plasmid pUC18-*copB* provides *in trans* a very high dosage of the *copB* gene cloned under control of its own constitutive promoter. When *S. Typhimurium* cells were transformed with pACYC184, similar transformation efficiencies were obtained irrespective of the presence of CopB in the recipient bacteria (Figure 2A). Also, the frequency of transformation of *S. Typhimurium* cells lacking CopB with pKN1562 was basically the same as with the p15A-derivative plasmid (Figure 2A). Contrarily, the presence of CopB most severely impaired the efficiency of transformation with the cognate plasmid containing the R1-replicon, as no transformant colonies appeared within 24 h of incubation of the plates (Figure 2A). The drastic and specific effect of CopB on the establishment of the R1 replicon was also observed in *E. coli*, where the presence of resident pUC18-*copB* reduced by more than two orders of magnitude the number of colonies transformed with pKN162, without affecting the efficiency of transformation with pACYC184 (Figure 2B).

With respect to the pMV158 system, its CopG repressor protein was also shown to significantly impair the establishment of the plasmid, although a differential effect was observed between *S. pneumoniae* and *S. aureus* (Figure 3). In these assays, resident plasmids pC194, pCGA30, pCGA3n, and pCGA3 provided no *copG*, inactive *copG* and medium and high dosages of the autoregulated active *copG* gene, respectively (see Table 2). The specific effect of CopG on the establishment of pMV158 in a new cell was analyzed by comparing the efficiency with which recipient strains containing each of these resident plasmids were transformed in parallel with the pMV158-derivative plasmid (pLS1) and with a pMV158-unrelated plasmid (pAMβ1 in *S. pneumoniae* and pT181-*cop608* in *S. aureus*).

In the pneumococcal host, the plasmids whose establishment was to be analyzed were introduced by natural transformation, a horizontal gene transfer mechanism that requires the development of a transient physiological property named competence. Many different factors have been shown to affect



**FIGURE 2 |** The presence of excess CopB in the recipient cell dramatically and specifically decreases the efficiency of transformation with the R1 replicon. The vertical bar graphs show the number of transformant colonies per ml that appeared after transforming either plasmid-free or pUC18-*copB*-carrying *S. Typhimurium* (A) and *E. coli* (B) cells with DNAs of plasmids pKN1562 (a mini-R1 derivative) or pACYC184 (harboring the R1-unrelated p15A replicon), both of which are compatible with the pUC18 replicon. The same volumes were plated for all transformed cultures; by plating this volume, 500–1,000 p15A-transformant colonies were counted. Transformant colonies were counted after incubation for 24 h at 30°C. The asterisk in (A) indicates the absence of transformants after transforming *S. Typhimurium* SL1344 carrying pUC18-*copB* with pKN1562. The ratio (Q) between the number of transformants per ml obtained with pKN1562 and that obtained with pACYC184 is indicated in the graphs on the right of the corresponding vertical bars.



**FIGURE 3 |** The presence of CopG in the recipient cell decreases the frequency or the velocity of appearance of transformants for the pMV158 replicon in a selective and dosage-dependent manner. (A) The vertical bar graph shows the frequency of transformants after transforming *S. pneumoniae* harboring different plasmids with pLS1 (pMV158 replicon) or pAMβ1 (pMV158-unrelated replicon). The resident plasmids provided no *copG* (pC194), inactive *copG* (pCGA30), and medium and high dosages of active *copG* gene (pCGA3n and pCGA3, respectively). The ratio (Q) between the frequency of transformants colonies obtained with pLS1 and that obtained with pAMβ1, counted after 60 h of incubation at 37°C, is indicated on the top of the corresponding vertical bars. (B) *S. aureus* cells, harboring the same set of plasmids as described in (A), were transformed with plasmids containing the replicon of either pMV158 or pT181. Vertical bars represent the frequency of transformants colonies counted after 24 h or 60 h of incubation at 37°C. The down facing blue arrow symbol indicates that the pMV158 replicon is only present as a plasmid co-integrate in the transformants. The ratio (Q) between the frequency of transformants colonies obtained with pLS1 and that obtained with pT181 *copG*608, counted after 60 h of incubation at 37°C, is indicated on the top of the corresponding vertical bars. The asterisk in (B) indicates the absence of transformant colonies appeared within 24 h of incubation. The data presented in this figure summarize the results obtained in typical transformation experiments of *S. pneumoniae* and *S. aureus*. Two additional transformation experiments were performed for each species and the results with respect to the inhibitory effect of CopG on the transformation with the pMV158 replicon were similar to those shown here.

competence (Attaiech et al., 2015), and we actually observed quite different chromosomal transformation frequencies in the various strains used. Namely, the strain harboring pCGA3 showed the highest transformation frequency, followed by the strain containing pCGA3n, next that harboring pCGA30, and finally the strain with the pC194 vector exhibited the lowest competence

level (see Material and Methods). The same qualitative trends were observed when analyzing the efficiencies of transformation of the various strains with plasmid pAMβ1 (whose replicon is not repressed by CopG), although in this case quantitatively larger differences were observed among them (Figure 3A). In order to normalize the frequencies of transformation with the

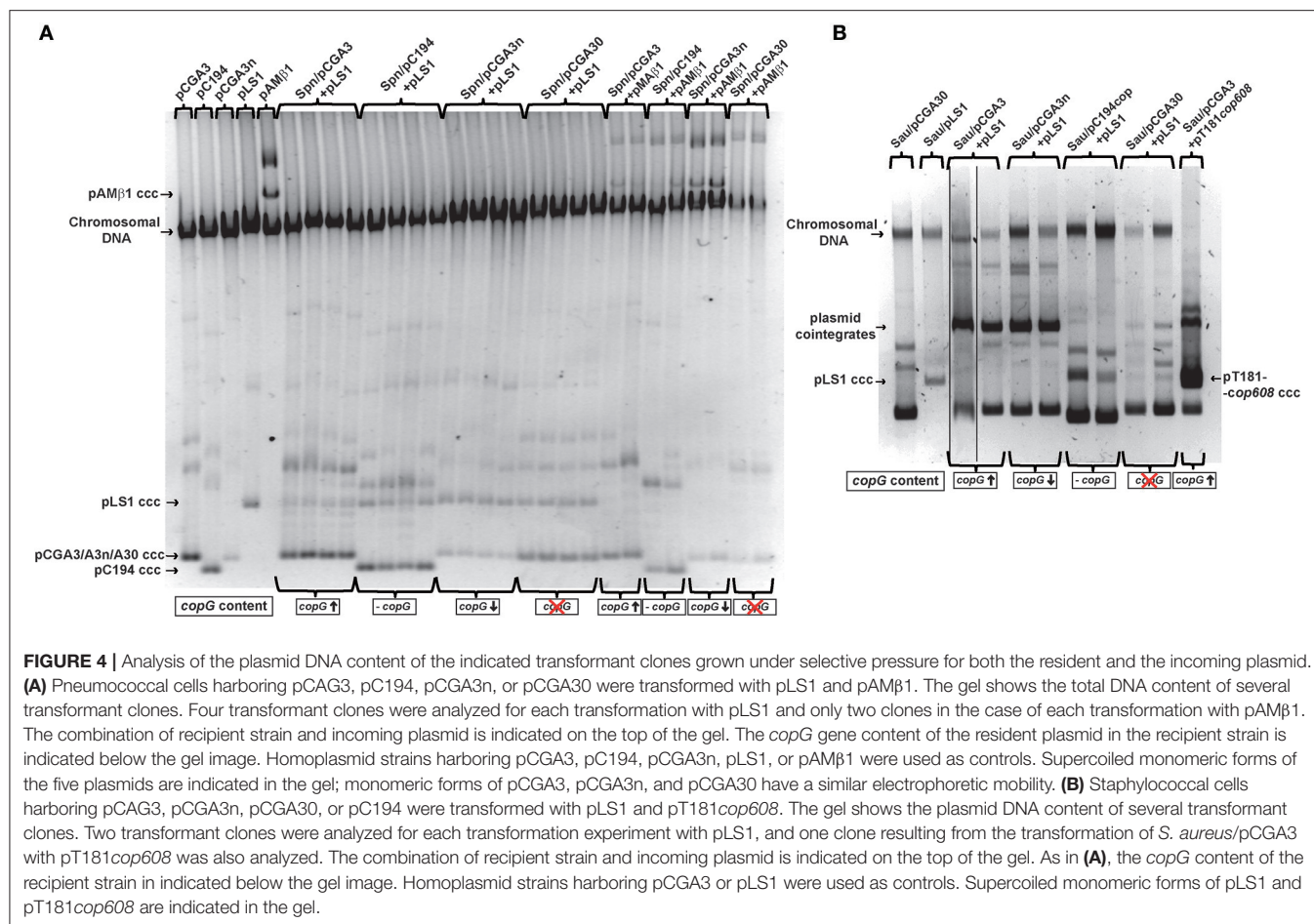


pMV158 derivative with respect to the level of competence for plasmid transfer, the ratio (Q) between the transformation efficiencies with the pMV158 derivative and with pAM $\beta$ 1 was used as parameter. This allowed us to evaluate the specific effect of CopG on the establishment of its cognate replicon. Since the pAM $\beta$ 1 transformants grew slowly, total transformant colonies were only counted after 60 h of incubation at 37°C, regardless of which plasmid was transferred into the recipient cells. When the void pC194 vector was the resident plasmid, pLS1 yielded 2.5-fold more transformants than pAM $\beta$ 1. This ratio increased to 24 when the recipient cells contained high-copy-number plasmid pCGA30 providing an inactive *copG* gene. The observed increase in the transformation efficiency is most likely due to facilitation of plasmid transformation by the existence of homology between the plasmid and the genome of the recipient cell, which in this case arises from the presence of a *copG* fragment in both the resident and the entering plasmid. The phenomenon of facilitation has only been reported to occur in natural transformation systems where donor DNA enters the competent cells in a linear ssDNA form (López et al., 1982). In spite of the potential facilitation of the transfer of the pMV158 derivative due to the presence of the entire *copG* gene in the recipient cells, the ratio of pLS1 to pAM $\beta$ 1 transformant colonies was reduced to 0.3 and to 0.02 when the resident plasmid provided, respectively, medium (pCGA3n) and high (pCGA3) dosages of active *copG* (Figure 3A). Two other independent lots of competent cells of the various strains were transformed in parallel with pLS1 and pAM $\beta$ 1, and similar Q ratios were obtained in both cases. These results show that the presence of CopG in the recipient cell severely and specifically impairs the success of the establishment of the cognate pMV158 replicon by decreasing the frequency of transformation, as we have also shown to be the case with CopB of the R1 system in both *E. coli* and *S. Typhimurium* (Figure 2).

In the staphylococcal host (Figure 3B), slightly different electrotransformation efficiencies were observed for the various strains, although all pT181-*cop608* transformant colonies appeared within the first 24 h of incubation irrespective of the presence or absence of CopG in the recipient cells. This was also the case for the pLS1 transformant colonies provided that the recipient strains lacked CopG (i.e., when the resident plasmids were pC194 or pCGA30). In contrast, virtually all pLS1 transformant colonies of the recipient strains carrying medium or high dosages of active *copG* could only be detected after 24 h of incubation, and thus they were counted at 60 h after plating. Despite the delay caused by CopG in the growth of the staphylococcal cells transformed with the pMV158 replicon, the ratio of pLS1 to pT181-*cop608* transformant colonies after 60-h incubation was close to 1, regardless of the plasmid resident in the recipient strain. Two more *S. aureus* transformation experiments were performed using the frozen stocks of electrocompetent cells. Although in these experiments the efficiencies of transformation with either plasmid were lower than when freshly-prepared cells were used, the Q ratios remained near constant. The mean Q value for all the strains was 1.1. Analysis of variance (ANOVA) of all experimental Q ratios indicated that there were no significant differences among groups, i.e., the same final frequency of pLS1 transformants was obtained irrespective of the presence of

CopG in the recipient cell. Hence, in *S. aureus* CopG seems to specifically impair transformation with the pMV158 replicon by decreasing the velocity of growth, but not the final frequency, of the transformants (Figure 3B).

To investigate the basis of the differential effect of CopG on the establishment of the pMV158 replicon in *S. pneumoniae* and *S. aureus*, we analyzed the plasmid content of various transformants of either species grown under selective pressure for both the resident and the incoming plasmid. This analysis was facilitated because the medium or high steady-state copy number of both resident and newly-acquired plasmids in these bacteria allows us to visualize the plasmid bands after electrophoretic separation and staining of DNA minipreps. In *S. pneumoniae*, total crude DNA extracts showed the presence of the expected resident and newly-acquired plasmids in all transformant clones (Figure 4A). When the co-resident plasmid provided no active *copG* (pC194 and pCGA30) or medium dosages of the active gene (pCGA3n), the steady-state copy number of pLS1 in the corresponding transformant clones remained the same as in the homoplasmid situation, whereas a ~35% decrease in the pLS1 copy number was observed if this plasmid coexisted with pCGA3, which provides high dosages of the active *copG* gene. A similar effect of the different dosages of *copG* supplied *in trans* on the steady-state copy number of the pMV158 replicon has been previously reported in transformant clones arising from the reverse transformations (i.e., when pneumococcal cells carrying pLS1 were transformed with the various pC194 derivatives) (del Solar et al., 1995). On the other hand, plasmid DNA minipreps of staphylococcal transformant clones (Figure 4B) only revealed the presence of the two expected plasmids when the resident plasmid was pC194 (no *copG*) or pCGA30 (inactive *copG*). In the latter case, however, an additional slight DNA band appeared that, according to restriction analysis, corresponded to a cointegrate generated by homologous recombination between the resident and the newly-acquired plasmids through the pMV158 DNA region cloned in the resident plasmid (not shown). When the resident plasmid carried an active *copG* gene (pCGA3 or pCGA3n), no separate pLS1 plasmid could be observed and the pMV158 replicon was only present as cointegrate (Figure 4B). It is worth noting that, unlike total DNA extracts from *S. pneumoniae*, the method used to extract the plasmid DNA from *S. aureus* gave random yields and thus could not be employed to estimate the plasmid copy number. Since CopG selectively inhibits the pMV158 replicon, generation of cointegrates with the resident plasmid allows incoming pLS1 to escape from replication inhibition and yet to reach the concentration required to render the host cell resistant to the antibiotic (Tc) with which transformants are selected. According to the results obtained, the exclusive presence of pLS1 as cointegrate occurs in all staphylococcal transformants where an active *copG* gene is provided by the resident plasmid (Figures 3B, 4B), whereas cointegrate formation does not seem to be the strategy used by the equivalent pneumococcal transformants (Figures 3A, 4A), even though cointegration between the incoming and resident plasmids can also occur in this bacterium (Figure S1). This distinct behavior might arise from differences in the frequency with which this Campbell-like recombination takes place in these two bacteria. Cointegrate formation seems to be ultimately



responsible for the delayed appearance of the staphylococcal transformant colonies.

To further analyze the role of the Cop proteins in the kinetics of plasmid repopulation, we chose the pMV158 replicon and its pneumococcal host as a model system because of a number of reasons. Importantly, compared to R1, the pMV158 replicon has a higher copy number in both the staphylococcal and the streptococcal host, and hence the amplitude of the plasmid amplification during the establishment phase replication is expected to be also greater, thus increasing the accuracy of the analysis. Moreover, the pneumococcal host of pMV158 was selected because we have a deep knowledge of it and we have set up a higher-frequency transformation protocol. And last but not least, no cointegrates that could mask repopulation of the pMV158 replicon have been observed in this system.

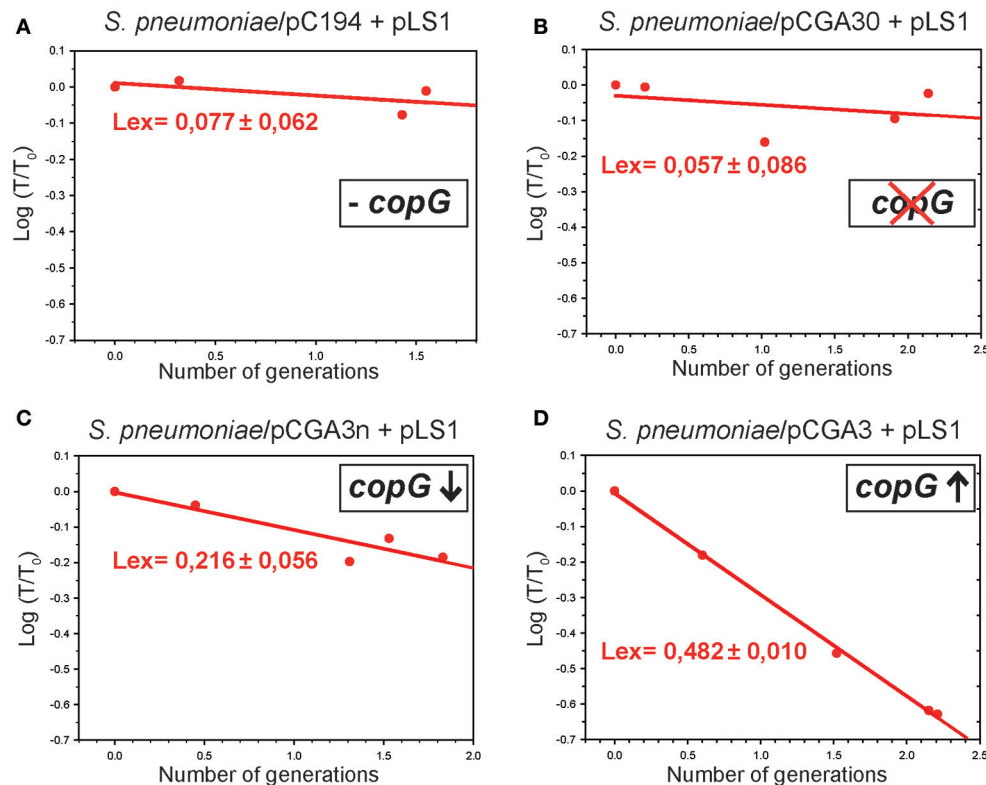
### The Presence of the pMV158 CopG Repressor Protein in the Recipient Cell Results in Segregational Instability of the Incoming pMV158 Replicon

As a first approach to study the effect of CopG-mediated transcriptional repression of the essential *repB* gene on the success and velocity of repopulation of the pMV158 replicon,

we tested the stability of inheritance of newly-acquired pLS1 in the transformant population of pneumococcal cells carrying null, medium, or high dosage of active *copG* (Figure 5). To this end, we analyzed the change in the numbers of total viable cells and transformants, as well as in the fraction of transformants, at various times after transformation. From these data, the experimental rate of loss of newly-acquired pLS1 from the transformants was calculated (see Material and Methods). The 3-layer *S. pneumoniae* plating method used in these assays (see Material and Methods) allowed the isolation of agar-embedded transformant c.f.u., so that plasmids could repopulate and express their antibiotic resistance determinant before selective pressure was applied.

When CopG-free recipient cells were employed (Figures 5A,B), no loss of pLS1 from the transformants could be inferred, although the fraction of transformants appeared to display a slight decreasing trend. This could be due to the pLS1 burden on the host, which for the steady-state plasmid concentration has been shown to cause an 8–9% increase in the cell doubling time (Hernández-Arriaga et al., 2012), so that the transformants slowly become overgrown by non-transformants.

When the plasmid resident in the recipient cell provided medium dosage of the active *copG* gene (Figure 5C), pLS1 showed unstable inheritance during division of the



**FIGURE 5 |** The presence of CopG in the recipient cell unstabilizes the inheritance of the incoming pMV158 replicon in a dosage-dependent manner. The graphs represent the changes in the fraction of transformants retaining the newly-acquired pMV158 replicon (pLS1) when growing for several generations in the absence of selective pressure for the plasmid. The recipient pneumococcal strains lacked *copG* (A), had a truncated version of *copG* (B), or harbored a medium (C) or high (D) dosage of the active *copG* gene. The experimental loss rate ( $L_{ex}$ ) of pLS1 was calculated from the slope of the linear regression model of the plot of the experimental values according to Equation (2) (red circles and lines).  $T_0$  and  $T$  are, respectively, the fractions of transformants *ab initio* and after  $n$  generations.

transformants, with a quite high loss rate per cell and generation ( $\sim 0.2$ ).

A near-maximal loss rate ( $\sim 0.5$ ) was determined for newly-acquired pLS1 when the recipient cells contained plasmid pCGA3, which provides high dosages of active *copG* gene (Figure 5D). This maximum segregational instability implies that most frequently pLS1 is inherited by only one of the two daughter cells resulting from division of the transformants. It should be mentioned that, in the pneumococcal host, pLS1 has been shown to be segregationally stable during the steady state, both in the homoplasmid situation (del Solar et al., 1987) and in the presence of recombinant plasmid pCGA3 (del Solar et al., 1995). Hence, plasmid loss in the transformants that harbor extra copies of the active *copG* gene can certainly be ascribed to failures in the establishment phase replication of the pMV158 derivative.

### The CopG Repressor Protein Impairs pMV158 Repopulation in the Transformants by Decreasing the Plasmid Replication Rate

As a further step toward the characterization of the role of CopG during the establishment phase replication of the

pMV158 replicon, we have analyzed the kinetics of repopulation of plasmid pLS1*cop7*, a copy-up derivative of pLS1, in pneumococcal cells carrying high dosages of either active or inactive *copG* gene. Compared to pLS1, pLS1*cop7* has a single-point mutation in the *copG* gene, thus encoding a defective CopG protein that leads to a 5-fold increase in the plasmid copy number (del Solar et al., 1990, 1995). The use of pLS1*cop7* in these assays ensures that the effects observed when the resident plasmid carries an active *copG* gene arise from the CopG protein present in the recipient cells and not from that encoded by the incoming plasmid, and also allows determining the replicon repopulation kinetics in the absence of any CopG. Moreover, the use of this copy-up derivative of pMV158 was expected to increase the amplitude of the replicative amplification in the case that repopulation occurred.

Plasmid stability assays showed that pLS1*cop7* was inherited rather stably (with no plasmid loss being inferred) when the recipient cells lacked CopG (Figure 6B). In fact, the fraction of transformants was kept almost constant along several generations of cell growth in the absence of selection for the incoming plasmid (Table 4). In contrast, pLS1*cop7* was lost from the CopG-containing transformants at about the maximum possible rate (0.5; Figure 6A), as can be inferred from the transformant

fraction being inversely proportional to the number of total cells (**Table 4**). The results of the segregational stability of newly-acquired pLS1*cop7* coincided with those obtained with incoming pLS1 (**Figure 5**), thus showing that the amount of CopG provided by the plasmid resident in the recipient cells suffices to severely impair repopulation of the pMV158 replicon.

The issue of the impact of CopG on the kinetics of pLS1*cop7* repopulation was first addressed by qPCR assays aimed at quantifying the variation of the number of copies of the incoming plasmid relative to the chromosome during the growth of the total bacterial population in the absence of Tc (**Figures 6C,D**). When the recipient cells contained resident pCGA30 and hence no functional CopG was provided, an abrupt decline (from  $\sim 1$  to  $\sim 0.04$ ) in the relative copy number of the plasmid amplicon within the total population was observed during the first 30 min of bacterial growth. This decrease was followed by a slower increase in the number of copies of the plasmid amplicon until a value of  $\sim 0.32$  was reached after 150 min (**Table 5B** and **Figure 6D**). On its turn, a smaller decrease (from  $\sim 0.07$  to  $\sim 0.01$ ) followed by a rather constant relative copy number of the plasmid amplicon was observed when the recipient cells provided CopG (**Table 5A** and **Figure 6C**). This transient high concentration of the plasmid specific amplicon that decayed very rapidly in the absence of cell division was most unlikely to correspond to intact plasmid molecules, and could rather reflect the features of the mechanism of natural transformation in *S. pneumoniae*. It is worth mentioning that donor plasmid DNA enters the pneumococcal cell as ssDNA segments of both strands, and that two or more fragments of the opposite strands must anneal through overlapping regions at their ends to generate a circular plasmid form with partial dsDNA regions. This is followed by DNA synthesis to reconstruct the intact plasmid molecule. It should also be noted that, in order to have a high frequency of transformation, we added 300–500 plasmid DNA molecules per bacterial cell. Thus, a number of ssDNA molecules that have entered the cell will harbor the plasmid amplicon to be amplified in the qPCR assay, although most of them will never render a reconstructed plasmid molecule and will be degraded by the cellular nucleases instead. The actual content in intact plasmid molecules was then analyzed by iPCR employing gDNA and two divergent primers that specifically annealed to DNA sequences within the pMV158 replicon (**Figure 6G**). Immediately after transformation with pLS1*cop7*, while the relative amount of the qPCR-detected pMV158 amplicon in the CopG-free recipient cells was maximal ( $t = 0$ ; **Figure 6D**), iPCR amplification rendered the faintest band of specific full-length plasmid DNA (**Figure 6F**). A similar result was observed when the CopG-containing cells were transformed with pLS1*cop7*, although in this case the initial amount of the pMV158 amplicon was much lower (**Figures 6C,E**). The observed discrepancy between the initial relative amounts of pMV158 amplicon and intact pLS1*cop7* DNA molecules led us to conclude that, in fact, most of the amplicons to be amplified in the qPCR assays were not carried on reconstituted plasmids but on smaller DNA fragments. Hence, the relative copy number of the donor plasmid shortly after entrance is better evaluated from the iPCR assays. At longer times after transformation (30 min and

further), when the DNA fragments carrying the amplicon would have declined, an almost perfect match was observed between the intracellular amplifications of the pLS1*cop7* amplicon (evaluated by qPCR) and of the intact plasmid molecules (evaluated by iPCR). This match was observed irrespective of whether the transformants carried or lacked CopG (**Tables 5A,B**; see also the Discussion). This coincidence allowed us to determine the rate of pLS1*cop7* replication ( $R$ , defined as the ratio of plasmid to gDNA duplications, and calculated according to Equation 10) along the time, based on the iPCR data of the plasmid amplification which, unlike the qPCR data, were unaffected by the presence of DNA fragments carrying the pMV158 amplicon.

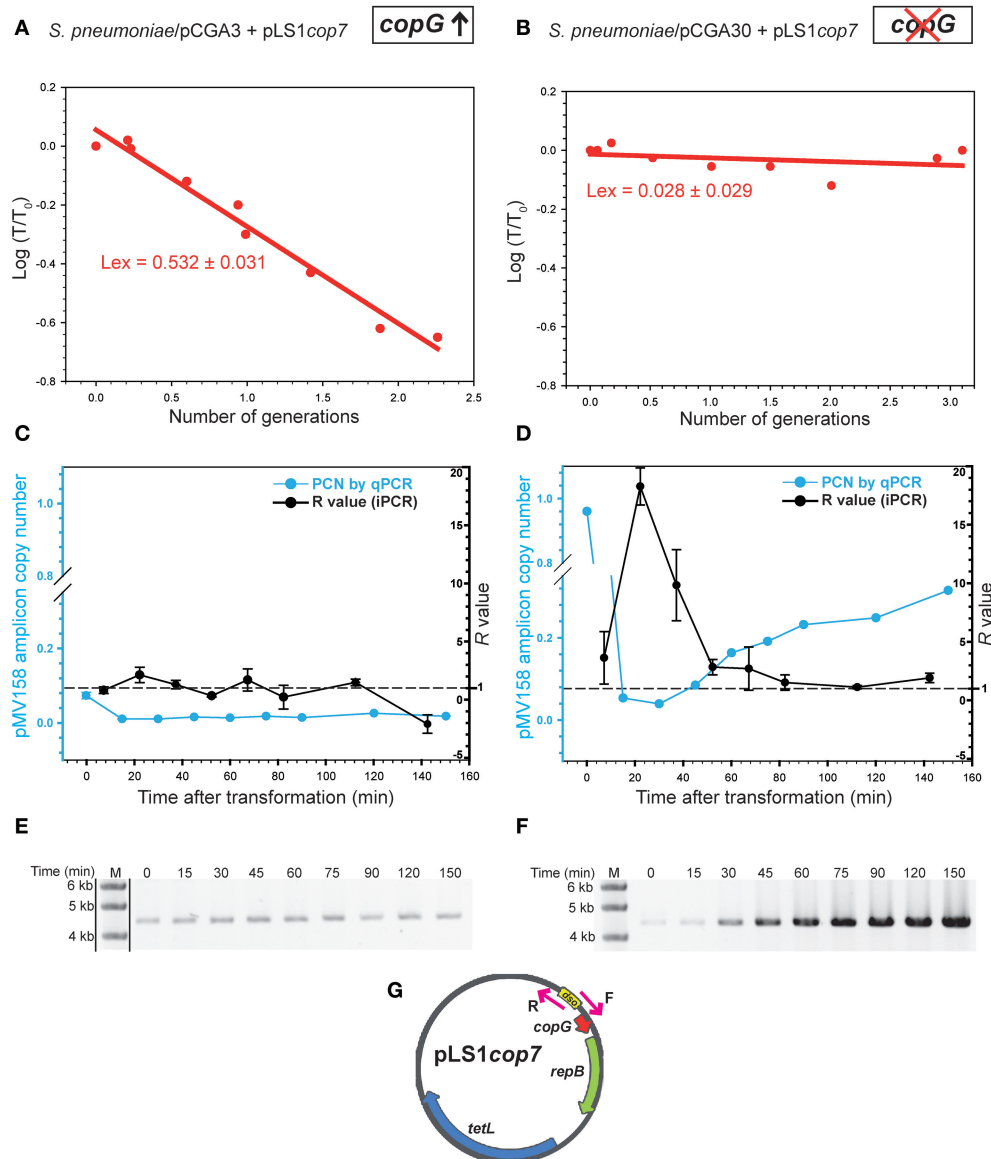
Although varied slightly along the time, and even reached a value of  $\sim 2$  in the interval between 15 and 30 min after transformation, the replication rate ( $R$ ) of pLS1*cop7* in the transformants carrying CopG fluctuated around 1 (**Table 5A** and **Figure 6C**). An  $R$  value of  $\sim 1$ , which was also inferred from the ratio between pDNA and gDNA total duplications ( $R = 1.1$ ; **Table 5A**), implies that the plasmid replicated at the same average velocity as the gDNA during the time interval analyzed and, hence, that it failed to repopulate.

When no CopG was present, an overall  $R$  value of  $\sim 2.8$  (the ratio of pDNA to gDNA total duplications; **Table 5B**) was found for incoming pLS1*cop7* during the entire 150-min interval analyzed, although the plasmid replication rate varied substantially along the time (**Figure 6D** and **Table 5B**). Repopulation of pLS1*cop7* mainly occurred during the first 45 min after completion of transformation, with a peak in the plasmid replication rate in the interval between 15 and 30 min (**Figure 6D** and **Table 5B**). Afterwards, the plasmid replication rate decreased asymptotically to 1, indicating that some repopulation still occurred at these longer times. There seemed to be also a small increase in the duplication rate of the pMV158 replicon when bacterial growth slowed down (in the 120–150 min interval), which would be also reflected in an increase of the plasmid copy number (**Figure 6D** and **Table 5B**). Whether this could indicate an increase of the pMV158 copy number during the stationary growth phase of its host is a potential matter for future investigation.

## DISCUSSION

In this work we have analyzed for the first time the role played by the Cop regulatory loops of R1 and pMV158 in plasmid establishment. The establishment phase replication, which amplifies the plasmid from an initially low concentration to the steady-state copy number, is a crucial process in the biology of naturally transferable plasmids that may importantly affect the success of their colonization and spreading. Nevertheless, plasmid replication during the establishment phase has been much less studied than the steady-state replication. In a pioneer work by Highlander and Novick (1987) the kinetics of repopulation of various pT181 derivatives that carried or lacked a functional asRNA control system were analyzed in *S. aureus* by determining the replication rates and copy numbers of the plasmids after radioactive *in vivo* labeling of total gDNA. A bit





**FIGURE 6 |** The presence of CopG in the recipient cell impairs repopulation of the pMV158 replicon by decreasing the plasmid replication rate. Changes in the fraction of transformants retaining newly-acquired pLS1cop7 (a copy-up pMV158 derivative) were analyzed in pneumococcal strains that either harbor a high dosage of active *copG* (**A**) or a truncated version of *copG* (**B**). The experimental loss rate ( $L_{ex}$ ) of pLS1cop7 was calculated from the slope of the linear regression model of the plot of the experimental values according to Equation (2) (red circles and lines).  $T_0$  and  $T$  are, respectively, the fractions of transformants *ab initio* and after  $n$  generations. The graphs in (**C**, **D**) show the impact of CopG on the kinetics of pLS1cop7 repopulation. A qPCR approach was used to calculate the variation in the copy number of a specific amplicon of the incoming pLS1cop7 plasmid relative to the chromosome during the growth of the total bacterial population containing pCGA3 (**C**) or pCGA30 (**D**) as resident plasmid for 150 min after transformation (left y-axis; blue circle and lines). The pLS1cop7 replication rate ( $R$  value), defined as the ratio of plasmid to gDNA duplications, was calculated at different time intervals following transformation of pneumococcal cells harboring pCGA3 (**C**) or pCGA30 (**D**). Determination of  $R$  was based on the iPCR data of the *in vivo* plasmid amplification (black circles and lines) and its value (right y-axis) was calculated according to Equation (10). Discontinuous horizontal line in graphs of (**C**, **D**) denotes an  $R$  value of 1, which characterizes the steady-state plasmid replication. The mean (symbols) and standard deviation (error bars) of all the experimental points in the graphs of (**C**, **D**) are displayed. Panels (**E**, **F**) show the iPCR analysis of the gDNA samples obtained at the indicated times after transformation of pneumococcal cells carrying pCGA3 and pCGA30, respectively, with pLS1cop7. iPCR assays were carried out by using a pair of divergent primers specific for the pMV158 replicon (Table 3 and **G**) and the Phusion DNA polymerase. Lane M, DNA molecular weight standard (NZYDNA ladder III; NZYTECH). Note that lanes M are the same in (**E**, **F**) because, in fact, both images of these panels arise from the same gel. Dividing lines in (**E**) indicate grouping of different parts of the same gel. The original image of the gel used for (**E**, **F**) composition is shown in Figure S2. A schematic representation of pLS1cop7 displaying the plasmid regions complementary to the divergent primers is shown in (**G**). Genes *copG*, *repB*, and *tetL*, as well as the *dso* region, are indicated.

**TABLE 4 |** Total number of cells, number of generations, and % of transformants at different times after transformation during the growth of the indicated strains in the absence of selective pressure.

Time after transformation (min)	<i>S. pneumoniae</i> /pCGA3 + pLS1cop7			<i>S. pneumoniae</i> /pCGA30 + pLS1cop7		
	Total number of cell generations <sup>a</sup>	Total cell count (c.f.u./ml)	% of transformants	Total number of cell generations <sup>a</sup>	Total cell count (c.f.u./ml)	% of transformants
0	0	$3.48 \times 10^7$	0.89	0	$2.32 \times 10^7$	0.16
15	0.01	$3.52 \times 10^7$	0.94	0.07	$2.48 \times 10^7$	0.16
30	0.23	$4.11 \times 10^7$	0.87	0.17	$2.62 \times 10^7$	0.17
45	0.60	$5.27 \times 10^7$	0.66	0.52	$3.32 \times 10^7$	0.15
60	0.94	$6.72 \times 10^7$	0.56	1.01	$4.77 \times 10^7$	0.14
75	0.99	$6.94 \times 10^7$	0.45	1.50	$6.59 \times 10^7$	0.14
90	1.42	$9.32 \times 10^7$	0.33	2.01	$9.37 \times 10^7$	0.12
120	1.88	$12.85 \times 10^7$	0.21	2.89	$17.20 \times 10^7$	0.15
150	2.26	$16.80 \times 10^7$	0.19	3.10	$19.95 \times 10^7$	0.16

<sup>a</sup>Total number of cell generations was calculated according to Equation (3).

later, the establishment phase replication of ColE1 was studied by using Southern blot to determine the number of plasmids containing the plasmid replicon per *E. coli* cell as a function of time after infection (Merlin and Polisky, 1993). It was found that a plasmid containing an up mutation in the RNA II primer promoter replicated at a 15-fold faster rate than the wild type, thus early highlighting the importance of rapid synthesis of the essential RNA II in ColE1 plasmid establishment. So far, however, the scarce characterized examples of repopulation have missed out the analysis of the effect on plasmid establishment of transcriptional repressors that either exert an auxiliary role or act synergistically with an antisense RNA in controlling the steady-state replication. As a first approach to address this analysis, we have investigated whether the presence of the R1 or pMV158 Cop repressor protein in the recipient cell affects the frequency and velocity of appearance of the plasmid transformant colonies. Moreover, by taking pMV158 and its pneumococcal host as a model system, we have developed a new approach for evaluating the kinetics of plasmid repopulation that is based on the estimation of the plasmid loss rate in transformants and on the use of non-radioactive highly sensitive qPCR and iPCR methods.

The results of the frequencies of transformation of different bacterial species with the R1 or the pMV158 replicons show that, when supplied in the recipient cell, CopB from R1 and CopG from pMV158 severely and selectively impair the establishment of their cognate plasmids. Actually, a dramatic decrease in the efficiency of transformation of *E. coli* or *S. Typhimurium* with the R1 replicon, and of *S. pneumoniae* with the pMV158 replicon can be observed when the cognate Cop repressor is present in the recipient cells (Figures 2, 3A). On its turn, in *S. aureus* cointegration with the resident plasmid allows the incoming pMV158 derivative to overcome the CopG-mediated inhibition of its replicon, so that only a delay in the appearance of the transformant colonies, but not a decrease in the final frequency of them, is observed in this bacterium. As seen for the pMV158 replicon/pneumococcal host system, inhibition of

plasmid establishment requires the presence of an active *copG* gene and depends on the dosage of this gene (Figure 3A). Since the R1 and pMV158 Cop proteins repress transcription of the respective plasmid *rep* gene from a strong promoter, the requirement of fully unrepressed expression of the essential *rep* genes for the successful establishment of these replicons can be inferred from the results of the transformation experiments performed in this work. The coincidence between the impairing effects of preexisting CopB and CopG on the efficiency of establishment of their cognate replicon in a new host leads us to think that the Cop repressor-mediated blockage of plasmid repopulation observed in the pMV158/pneumococcal host system can be extrapolated to R1 entering its *Enterobacteriaceae* host.

As shown in Figure 7, unsuccessful repopulation of the incoming plasmid may lead to unstable inheritance of the plasmid during division of the transformed cells. The rates at which the pMV158 derivatives (wild-type or copy-up mutant plasmids) are lost during the culture of pneumococcal transformants carrying or lacking a high dosage of autoregulated active *copG* gene match quite well the two extreme theoretical cases of failure or immediate plasmid repopulation, respectively (Figure 7). In fact, a loss rate close to the maximum value (0.5) was found in the first case whereas no significant loss was observed in the second one (Figures 5A,B,D, 6A,B). When a medium dosage of autoregulated *copG* gene is provided by the recipient cell, an intermediate rate of loss of the pMV158 replicon was observed (Figure 5C). Actually, these results indicate that the presence of a functional *copG* gene in the recipient cell causes a dosage-dependent impairment in the pMV158-replicon repopulation that leads to the unstable inheritance of the underpopulated plasmid.

The kinetics of repopulation of the pMV158 replicon, and the effect of the CopG repressor protein on it, have been studied by qPCR and iPCR using total gDNA, prepared at different times after transformation, as template. Irrespective of whether the

**TABLE 5 |** Kinetics of genomic and plasmid DNA replication during plasmid establishment.

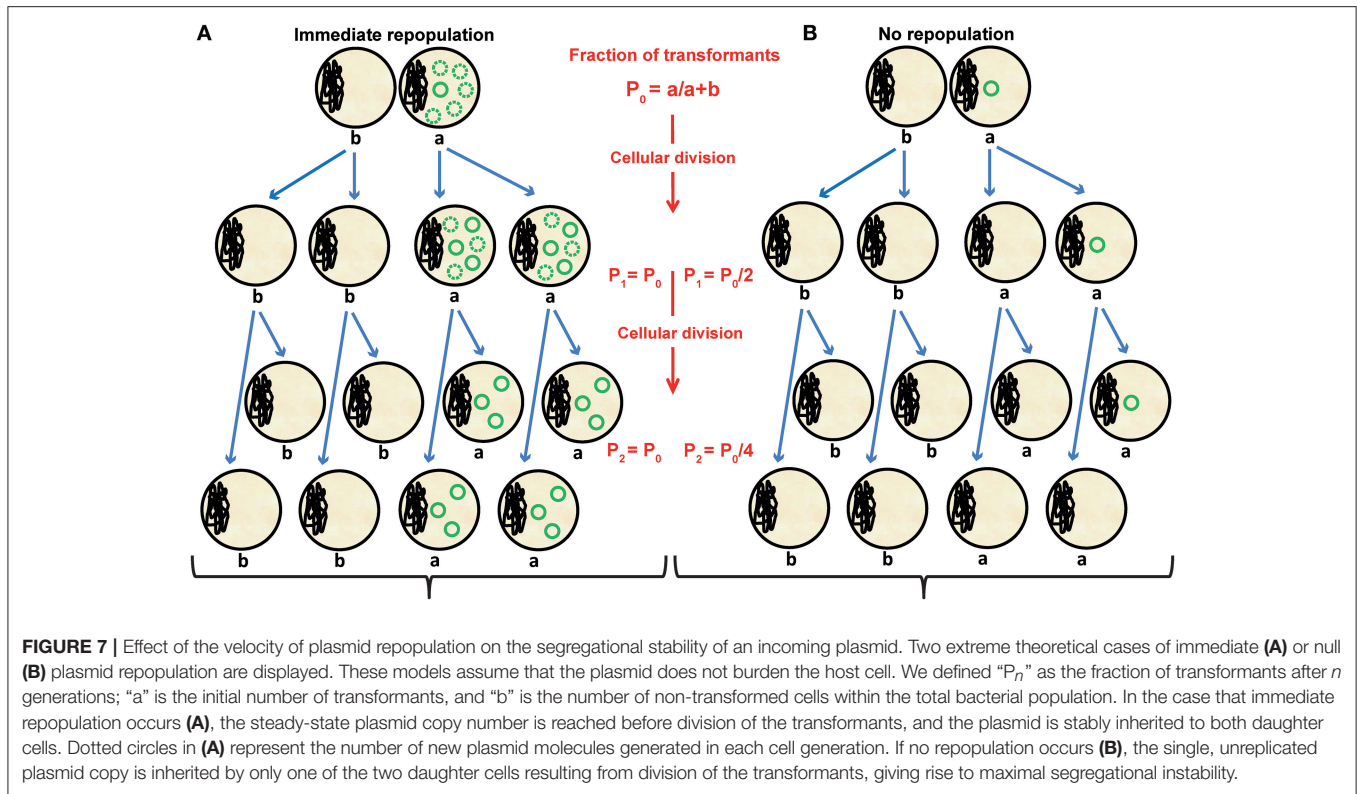
<i>S. pneumoniae</i> /pCGA3 + pLS1cop7								
Time after transformation (min)	Total number of cell generations <sup>a</sup>	Copy number of the plasmid amplicon (qPCR)	Time interval	Amplification of plasmid amplicon <sup>b,d</sup> (qPCR)	Amplification of plasmid molecules <sup>c,d</sup> (iPCR)	R <sup>e</sup>	gDNA duplications	pDNA duplications
<b>A</b>								
0	0	0.074 ± 0.009						
15	0.01	0.011 ± 0.002	0–15	0.2	1.0	0.8 ± 0.3	0.21	0.17
30	0.23	0.011 ± 0.001	15–30	1.0	1.3	2.1 ± 0.7	0.28	0.60
45	0.60	0.016 ± 0.001	30–45	1.4	1.1	1.3 ± 0.4	0.60	0.80
60	0.94	0.014 ± 0.001	45–60	0.9	0.9	0.3 ± 0.1	0.35	0.12
75	0.99	0.018 ± 0.001	60–75	1.3	1.0	1.7 ± 1.0	0.16	0.27
90	1.42	0.015 ± 0.001	75–90	0.8	0.9	0.2 ± 1.0	0.41	0.09
120	1.88	0.026 ± 0.001	90–120	1.7	1.5	1.5 ± 0.3	0.96	1.43
150	2.26	0.018 ± 0.001	120–150	0.7	0.8	−2.1 ± 0.8	0.06	−0.10
							Total = 3.03	Total = 3.36
<i>S. pneumoniae</i> /pCGA30 + pLS1cop7								
Time after transformation (min)	Total number of cell generations <sup>a</sup>	Copy number of the plasmid amplicon (qPCR)	Time interval	Amplification of plasmid amplicon <sup>b,d</sup> (qPCR)	Amplification of plasmid molecules <sup>c,d</sup> (iPCR)	R <sup>e</sup>	gDNA duplications	pDNA duplications
<b>B</b>								
0	0	0.960 ± 0.010						
15	0.07	0.054 ± 0.020	0–15	0.1	1.4	3.0 ± 1.7	0.24	0.72
30	0.17	0.040 ± 0.001	15–30	0.7	5.1	14.1 ± 1.2	0.18	2.54
45	0.52	0.085 ± 0.013	30–45	2.1	2.3	7.7 ± 2.3	0.18	1.39
60	1.01	0.164 ± 0.021	45–60	1.9	1.7	2.4 ± 0.5	0.54	1.30
75	1.50	0.191 ± 0.025	60–75	1.2	1.2	2.3 ± 1.4	0.20	0.46
90	2.01	0.232 ± 0.023	75–90	1.2	1.1	1.4 ± 0.5	0.34	0.48
120	2.89	0.249 ± 0.018	90–120	1.1	1.1	1.1 ± 0.1	1.10	1.22
150	3.10	0.315 ± 0.025	120–150	1.3	1.2	1.7 ± 0.3	0.38	0.65
							Total = 3.16	Total = 8.76

<sup>a</sup>Total number of cell generations was calculated according to Equation (3). <sup>b,c</sup>The factor by which the <sup>b</sup>plasmid amplicon or the <sup>c</sup>plasmid molecules copy number is increased in the indicated time interval. <sup>d</sup>Amplification factor for a certain time interval can be calculated as the product of the amplification factors of the reference time intervals. <sup>e</sup>R value was calculated according to Equation (10).

pneumococcal cells harbor or not the *copG* gene, replication of the gDNA (consisting mainly of chromosomal DNA) appears to begin earlier than cellular division after the 30 to 37°C shift that follows the transformation step (Table 5). The estimated number of total gDNA duplications in the 150-min interval is about three for both kinds of recipient cells, a value that almost equals the number of total cell generations in the case of the recipient cells containing the inactive *copG* gene, whereas it is somewhat higher than the value obtained when the recipient cells contain a functional *copG* gene (Table 5).

In the absence of CopG, the highest repopulation rate occurs during the first 45 min after the entrance of the plasmid DNA in the cell, with a peak in the interval between 15 and 30 min (Figure 6D and Table 5B). According to the results obtained by

qPCR (Table 5B), and taking also into account the fraction of transformants (Table 4), the average copy number of pLS1cop7 in the transformant cells is estimated to be about 200 per chromosome equivalent at 150 min after plasmid transfer. This value coincides with the reported steady-state copy number of the copy-up pMV158 derivative (Table 2), indicating that repopulation has been accomplished by this time. Moreover, no overshoot of the steady-state copy number was observed and instead this value was asymptotically approached (Figures 6D,F). Overall, about nine duplications of the pMV158 derivative took place within the transformants harboring an inactive *copG* gene in the 150-min time interval during which the gDNA duplicated around three times (Table 5B). This represents a 50- to 60-fold relative amplification of the plasmid DNA (see Equation 8),



which can also be determined from the amplification factor of the plasmid copy number (Table 5B).

In the presence of CopG, the relative average plasmid copy number in the total bacterial population is kept near constant, as shown by the results from both the qPCR and the iPCR assays (Figure 6C,E and Table 5A). This is also consistent with an overall plasmid replication rate ( $R$ ) of 1.1 along the entire 150-min time period after transformation (Table 5A), which means that, on average, the incoming pMV158 replicon underwent about the same number of duplications as the gDNA. An  $R$  value of around 1, which in fact characterizes the steady-state replication of any plasmid, demonstrates the unsuccessful repopulation of the pMV158 replicon when CopG is present in the recipient cell. However, based on the maximum possible value of plasmid loss rate that is observed (Figure 6A), a total absence of plasmid replication ( $R \sim 0$  at any time interval) should have been expected instead (Figure 7B). We could envisage two potential explanations for this. One possibility is that in most cells the plasmid replicates at approximately the same rate as the chromosome, thus giving rise to cells containing at least two plasmid copies at division; if so, the observed loss rate would require that the two sister plasmid molecules tend to segregate together into the same daughter cell. An alternative explanation is that plasmid replication manages to evade inhibition by CopG in a small fraction of the transformants so that repopulation to the steady-state plasmid copy number is achieved only in these cells, and hence the maximum value of plasmid loss rate (0.5) would not be significantly altered. Be that as it may, according to the qPCR assays the relative average plasmid copy number in the total bacterial population increases  $\sim 1.65$ -fold during the time

interval between 15 min (when the plasmid DNA fragments are considered to be negligible) and 150 min after transformation (Table 5A). When corrected for the fraction of transformants (Table 4), the average plasmid copy number per chromosome in the sub-population of transformed cells is shown to increase from  $\sim 1.2$  to  $\sim 9.5$  between 15 min and 150 min after transformation, although at the latter time the plasmid copies might not be evenly distributed among the transformants, as discussed above. Moreover, once the transformant clones are selected for the incoming plasmid, pLS1cop7 seems to be “forced” to repopulate until reaching the steady-state relative copy number ( $\sim 33$ ) that is observed in the heteroplasmid situation (Figure S1).

It should be noted that we have obtained similar negative effects on plasmid establishment due to the presence of Cop proteins in the recipient cells when repopulation followed either of two distinct horizontal transfer mechanisms, namely a natural transformation or an electrically-induced transfer. Although the R1 and pMV158 derivatives used in this work did not allow the analysis of plasmid repopulation following conjugative transfer, the significance of fully unrepressed transcription of their essential *rep* genes can most likely be extrapolated to any process implying the establishment phase replication. Establishment phase repopulation is a pivotal step within the process of horizontal plasmid spread by which a unique (or very few) intact dsDNA molecule of the donor plasmid is amplified to the steady-state characteristic copy number. Depending on the transfer mechanism employed, one full-length circular dsDNA plasmid copy can enter directly the cell (electroporation), can be reconstructed from overlapping ssDNA fragments of both strands (pneumococcal natural transformation), or can be



generated by synthesis of the DNA strand complementary to the one that was transferred by conjugation. Nevertheless, plasmid repopulation (and its Cop-mediated inhibition) only depends on the replicon function and is independent of the mechanism used for plasmid transfer. In this sense, it is worth noting that similar replication kinetics were observed for the repopulation of pT181 derivatives in *S. aureus* irrespective of whether the initial low plasmid copy number was achieved by shutoff of the replication of thermosensitive mutants or by introduction of the plasmid DNA into the bacterial cells through high-frequency (50%) transduction (Highlander and Novick, 1987).

Carrying an emergency mechanism that enables strong expression of the essential *rep* gene allows rapid and successful repopulation, which can benefit especially those plasmids whose lifestyle includes colonization of new hosts. This is in fact the case, among others, of conjugative or mobilizable plasmids R1, pIP501, and pMV185, all of which harbor a strong promoter directing transcription of the *rep* gene that is subjected to repression by a plasmid-encoded Cop protein. Although a crucial role in plasmid establishment has been proposed for the Cop regulatory loops of these plasmids, no empirical demonstration of it had been reported so far (Nordström and Nordström, 1985; del Solar et al., 1990; del Solar and Espinosa, 2000; Olsson et al., 2004; Brantl, 2014). We think that the requisite, observed in the present work, of fully unrepressed *rep* transcription for the successful establishment of R1 and pMV158 can be extended to other plasmids encoding a similar Cop regulatory loop.

Understanding the role of the Cop regulatory loop that switches on/off transcription of the essential *rep* gene may help the design and development of new strategies to control spreading of undesirable plasmids among bacterial populations or to prevent transfer of a specific plasmid to a potential host in mating experiments with multiple plasmids.

## AUTHOR CONTRIBUTIONS

RD-O conceived the idea of bringing together the independent observations obtained with the R1 and pMV158 plasmid systems

to prepare a joining article about the role of Cop repressors on plasmid establishment. RD-O and IM-C designed and performed the transformation experiments with the R1 replicon. LL performed the transformation experiments with the pMV158 replicon. JR-M carried out the experiments of the kinetics of pMV158 repopulation and the analysis of plasmid stability in the transformants, wrote the Material and Methods section and also prepared most of the figures. GdS designed the experimental approach and the formulation of the kinetics of repopulation and plasmid stability analyses, participated in the experiments with the pMV158 system and wrote most of the manuscript. All the authors discussed the results and corrected the entire manuscript.

## FUNDING

This work was supported by the Spanish Ministry of Economy and Competitiveness (SMEC; grant AGL2015-65010-C3-1-R). We are also grateful to the members of the consortium “Interactivities between plasmid modules and bacterial chromosomes: a due visit” (to which RD-O, JR-M, and GdS belong; grant BIO2015-69085-REDC by SMEC) for fruitful discussions.

## ACKNOWLEDGMENTS

Thanks are due to members of our labs for helpful discussions, and particularly to Lidia de Tapia for her contribution to the *Salmonella* transformation experiments and to Sergio Barata, Javier Nicolas Garay Novillo and Diego García de la Morena for critical reading of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02367/full#supplementary-material>

## REFERENCES

- Appleyard, R. K. (1954). Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. *Genetics* 39, 440–452.
- Atlung, T., Christensen, B. B., and Hansen, F. G. (1999). Role of the rom protein in copy number control of plasmid pBR322 at different growth rates in *Escherichia coli* K-12. *Plasmid* 41, 110–119. doi: 10.1006/plas.1998.1386
- Attaiach, L., Minnen, A., Kjos, M., Gruber, S., and Veening, J. W. (2015). The ParB-*parS* chromosome segregation system modulates competence development in *Streptococcus pneumoniae*. *MBio* 6, e00662-15. doi: 10.1128/mBio.00662-15
- Augustin, J., and Götz, F. (1990). Transformation of *Staphylococcus epidermidis* and other staphylococcal species with plasmid DNA by electroporation. *FEMS Microbiol. Lett.* 54, 203–207. doi: 10.1111/j.1574-6968.1990.tb03997.x
- Boer, D. R., Ruiz-Masó, J. A., Rueda, M., Pethoukov, M., Svergun, D. I., Orozco, M., et al. (2016). Conformational plasticity of RepB, the replication initiator protein of promiscuous streptococcal plasmid pMV158. *Sci. Rep.* 6:20915. doi: 10.1038/srep20915
- Boer, D. R., Ruiz-Masó, J. A., López-Blanco, J. R., Blanco, A. G., Vives-Llàcer, M., Chacón, P., et al. (2009). Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. *EMBO J.* 28, 1666–1678. doi: 10.1038/emboj.2009.125
- Brantl, S. (2014). Plasmid replication control by antisense RNAs. *Microbiol. Spectr.* 2:PLAS-0001-2013. doi: 10.1128/microbiolspec.PLAS-0001-2013
- Burdett, V. (1980). Identification of tetracycline-resistant R-plasmids in *Streptococcus agalactiae* (group B). *Antimicrob. Agents Chemother.* 18, 753–760. doi: 10.1128/AAC.18.5.753
- Chang, A. C., and Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134, 1141–1156.
- Das, N., Valjavec-Gratian, M., Basuray, A. N., Fekete, R. A., Papp, P. P., Paulsson, J., et al. (2005). Multiple homeostatic mechanisms in the control of P1 plasmid replication. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2856–2861. doi: 10.1073/pnas.0409790102
- del Solar, G., Acebo, P., and Espinosa, M. (1995). Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II. *Mol. Microbiol.* 18, 913–924. doi: 10.1111/j.1365-2958.1995.1805.0913.x

- del Solar, G., and Espinosa, M. (2000). Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* 37, 492–500. doi: 10.1046/j.1365-2958.2000.02005.x
- del Solar, G., Moscoso, M., and Espinosa, M. (1993). *In vivo* definition of the functional origin of replication (*ori*+) of the promiscuous plasmid pLS1. *Mol. Gen. Genet.* 237, 65–72. doi: 10.1007/BF00282785
- del Solar, G. H., Pérez-Martín, J., and Espinosa, M. (1990). Plasmid pLS1-encoded RepA protein regulates transcription from *repAB* promoter by binding to a DNA sequence containing a 13-base pair symmetric element. *J. Biol. Chem.* 265, 12569–12575.
- del Solar, G. H., Puyet, A., and Espinosa, M. (1987). Initiation signals for the conversion of single stranded to double stranded DNA forms in the streptococcal plasmid pLS1. *Nucleic Acids Res.* 15, 5561–5580. doi: 10.1093/nar/15.14.5561
- Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* 16, 6127–6145. doi: 10.1093/nar/16.13.6127
- Gerdes, K., Larsen, J. E., and Molin, S. (1985). Stable inheritance of plasmid R1 requires two different loci. *J. Bacteriol.* 161, 292–298.
- Gomis-Rüth, F. X., Solà, M., Acebo, P., Párraga, A., Guasch, A., Eritja, R., et al. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.* 17, 7404–7415. doi: 10.1093/emboj/17.24.7404
- Hernández-Arriaga, A. M., Espinosa, M., and del Solar, G. (2012). Fitness of the pMV158 replicon in *Streptococcus pneumoniae*. *Plasmid* 67, 162–166. doi: 10.1016/j.plasmid.2012.01.005
- Hernández-Arriaga, A. M., Rubio-Lepe, T. S., Espinosa, M., and del Solar, G. (2009). Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes. *Nucleic Acids Res.* 37, 4799–4811. doi: 10.1093/nar/gkp503
- Highlander, S. K., and Novick, R. P. (1987). Plasmid repopulation kinetics in *Staphylococcus aureus*. *Plasmid* 17, 210–221. doi: 10.1016/0147-619X(87)90029-1
- Hoskins, J., Alborn, W. E. Jr., Arnold, J., Blaszczyk, L. C., Burgett, S., DeHoff, B. S., et al. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* 183, 5709–5717. doi: 10.1128/JB.183.19.5709-5717.2001
- Hoiseth, S. K., and Stocker, B. A. (1981). Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature* 291, 238–239. doi: 10.1038/291238a0
- Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. doi: 10.1093/bioinformatics/btm091
- Kreiswirth, B. N., Lofdahl, S., Betley, M. J., O'Reilly, M., Schlievert, P. M., Bergdoll, M. S., et al. (1983). The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* 305, 709–712. doi: 10.1038/305709a0
- Lacks, S. A., and Greenberg, B. (1977). Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J. Mol. Biol.* 114, 153–168. doi: 10.1016/0022-2836(77)90289-3
- Lacks, S. A., López, P., Greenberg, B., and Espinosa, M. (1986). Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. *J. Mol. Biol.* 192, 753–765. doi: 10.1016/0022-2836(86)90026-4
- Leblanc, D. J., and Lee, L. N. (1984). Physical and genetic analyses of streptococcal plasmid pAM beta 1 and cloning of its replication region. *J. Bacteriol.* 157, 445–453.
- López, P., Espinosa, M., Stassi, D. L., and Lacks, S. A. (1982). Facilitation of plasmid transfer in *Streptococcus pneumoniae* by chromosomal homology. *J. Bacteriol.* 150, 692–701.
- López-Aguilar, C., Romero-López, C., Espinosa, M., Berzal-Herranz, A., and del Solar, G. (2015). The 5'-tail of antisense RNAII of pMV158 plays a critical role in binding to the target mRNA and in translation inhibition of *repB*. *Front. Genet.* 6:225. doi: 10.3389/fgene.2015.00225
- López-Villarejo, J., Lobato-Márquez, D., and Díaz-Orejas, R. (2015). Coupling between the basic replicon and the *kis-kid* maintenance system of plasmid R1: modulation by *Kis* antitoxin levels and involvement in control of plasmid replication. *Toxins* 7, 478–492. doi: 10.3390/toxins7020478
- Merlin, S., and Polisky, B. (1993). Analysis of establishment phase replication of the plasmid ColE1. *J. Mol. Biol.* 230, 137–150. doi: 10.1006/jmbi.1993.1131
- Molin, S., Stougaard, P., Uhlin, B. E., Gustafsson, P., and Nordström, K. (1979). Clustering of genes involved in replication, copy number control, incompatibility, and stable maintenance of the resistance plasmid R1rd-19. *J. Bacteriol.* 138, 70–79.
- Nordström, K. (1990). Control of plasmid replication - how do DNA iterons set the replication frequency? *Cell* 63, 1121–1124. doi: 10.1016/0092-8674(90)90405-4
- Nordström, K. (2006). Plasmid R1-Replication and its control. *Plasmid* 55, 1–26. doi: 10.1016/j.plasmid.2005.07.002
- Nordström, K., Molin, S., and Light, J. (1984). Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system. *Plasmid* 12, 71–90. doi: 10.1016/0147-619X(84)90054-4
- Nordström, M., and Nordström, K. (1985). Control of replication of FII plasmids: comparison of the basic replicons and of the *copB* systems of plasmids R100 and R1. *Plasmid* 13, 81–87. doi: 10.1016/0147-619X(85)90060-5
- Olsson, J. A., Paulsson, J., and Nordström, K. (2004). Effect of the CopB auxiliary replication control system on stability of maintenance of Par+ plasmid R1. *J. Bacteriol.* 186, 207–211. doi: 10.1128/JB.186.1.207-211.2004
- Pimentel, B., Madine, M. A., and de la Cueva-Méndez, G. (2005). Kid cleaves specific mRNAs at UUACU sites to rescue the copy number of plasmid R1. *EMBO J.* 24, 3459–3469. doi: 10.1038/sj.emboj.7600815
- Providenti, M. A., O'Brien, J. M., Ewing, R. J., Paterson, E. S., and Smith, M. L. (2006). The copy-number of plasmids and other genetic elements can be determined by SYBR-Green-based quantitative real-time PCR. *J. Microbiol. Methods* 65, 476–487. doi: 10.1016/j.mimet.2005.09.007
- Rasooly, A., Wang, P. Z., and Novick, R. P. (1994). Replication-specific conversion of the *Staphylococcus aureus* pT181 initiator protein from an active homodimer to an inactive heterodimer. *EMBO J.* 13, 5245–5251.
- Riise, E., and Molin, S. (1986). Purification and characterization of the CopB replication control protein, and precise mapping of its target site in the R1 plasmid. *Plasmid* 15, 163–171. doi: 10.1016/0147-619X(86)90034-X
- Rosenfeld, R., and Grover, N. B. (1993). Control of mini-R1 plasmid replication: a computer simulation. *Plasmid* 29, 94–116. doi: 10.1006/plas.1993.1012
- Ruiz-Masó, J. A., López-Zumel, C., Menéndez, M., Espinosa, M., and del Solar, G. (2004). Structural features of the initiator of replication protein RepB encoded by the promiscuous plasmid pMV158. *Biochim. Biophys. Acta Proteins Proteomics* 1696, 113–119. doi: 10.1016/j.bbapap.2003.09.010
- Ruiz-Masó, J. A., Lurz, R., Espinosa, M., and del Solar, G. (2007). Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication. *Nucleic Acids Res.* 35, 1230–1244. doi: 10.1093/nar/gkl1099
- Ruiz-Masó, J., Machón, C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M., and del Solar, G. (2015). Plasmid rolling-circle replication. *Microbiol Spectr.* 3:PLAS-0035-2014. doi: 10.1128/microbiolspec.PLAS-0035-2014
- Stassi, D. L., López, P., Espinosa, M., and Lacks, S. A. (1981). Cloning of chromosomal genes in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* 78, 7028–7032. doi: 10.1073/pnas.78.11.7028
- Summers, D. K. (2009). *The Biology of Plasmids*. Oxford, UK: Blackwell Publishing Ltd. doi: 10.1002/9781444313741
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596
- Wagner, E. G., Altuvia, S., and Romby, P. (2002). Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.* 46, 361–398. doi: 10.1016/S0065-2660(02)46013-0

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ruiz-Masó, Luengo, Moreno-Córdoba, Díaz-Orejas and del Solar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Importance of the Expendable: Toxin–Antitoxin Genes in Plasmids and Chromosomes

Ramón Díaz-Orejas<sup>1</sup>, Manuel Espinosa<sup>1\*</sup> and Chew Chieng Yeo<sup>2\*</sup>

<sup>1</sup> Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, <sup>2</sup> Faculty of Medicine, Biomedical Research Centre, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

## OPEN ACCESS

### Edited by:

Johann Peter Gogarten,  
University of Connecticut,  
United States

### Reviewed by:

Francis Repolla,  
Institut National de la Recherche  
Agronomique (INRA), France  
Nikolai Ravin,  
Research Center for Biotechnology  
(RAS), Russia

### \*Correspondence:

Manuel Espinosa  
mespinosa@cib.csic.es  
Chew Chieng Yeo  
chewchieng@gmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 09 May 2017

Accepted: 24 July 2017

Published: 04 August 2017

### Citation:

Díaz-Orejas R, Espinosa M and  
Yeo CC (2017) The Importance of the  
Expendable: Toxin–Antitoxin Genes  
in Plasmids and Chromosomes.  
Front. Microbiol. 8:1479.  
doi: 10.3389/fmicb.2017.01479

Toxin–antitoxin (TA) genes were first reported in plasmids and were considered expendable genetic cassettes involved in the stable maintenance of the plasmid replicon by interfering with growth and/or viability of bacteria in which the plasmid was lost. TAs were later found in bacterial chromosomes and also in integrated mobile genetic elements; they were proposed to be involved in the bacterial response to stressful situations. At present, 100s of TAs have been identified and classified in up to six families (I to VI), with those belonging to the type II (constituted by two protein components) being the most studied. Based on well-characterized examples of several type II TAs, we discuss in this review that irrespective of their locations in plasmids or chromosomes, TAs functionally overlap as indicated by: (i) in both locations they can mediate the maintenance of genetic elements to which they are physical linked, and (ii) they can induce persistence or virulence in response to stress situations. Examples of functional confluences in homologous TA systems with different locations are also given. We also consider whether the physiological role of TAs is due to their genetic organization as operons or to their inherent properties, like the short lifespan of the antitoxin component.

**Keywords:** toxin–antitoxin, plasmids, post-segregational killing, genomic islands, chromosome, bacterial virulence, persistence

## INTRODUCTION

Toxin–antitoxin (TA) genes were initially discovered in two conjugative plasmids of *Escherichia coli*, F (Ogura and Hiraga, 1983) and R1 (Gerdes et al., 1986a; Bravo et al., 1988; Tsuchimoto et al., 1988), as cassette of two genes involved in stable maintenance of these plasmids; they were shown to participate in stable plasmid inheritance because they reduced either the viability or the growth of the cells that had lost the plasmid at the time of cell division (Bravo et al., 1988; Tsuchimoto et al., 1988). Killing of plasmid free segregants was termed post-segregational killing (PSK) (Gerdes et al., 1986b) and shown to be due to the decay of the more unstable antitoxin in plasmid-free cells and to the subsequent activation of the toxin in these cells. A role of plasmidic TAs in outcompeting compatible plasmids was later proposed (Cooper and Heinemann, 2000). TAs were also subsequently shown to be encoded by bacterial chromosomes and some of them integrated within mobile genetic elements (MGEs). One of the hypotheses to explain the presence and function of these TAs was that they participate in the response to stressful conditions

(Christensen and Gerdes, 2003). At present, 100s of TAs have been identified and classified, depending on the nature and activity of antitoxins, in up to six types (type I to type VI), reviewed in Page and Peti (2016). To these toxin–antitoxin pairs can also be added Type II restriction enzymes and their cognate methylases: similar to conventional TAs, restriction-modification systems can also be encoded in plasmids and enforce their maintenance by promoting PSK of plasmid-free cells (Yarmolinsky, 1995), although it must be emphasized that TA and RM complexes that have been characterized so far do not share an evolutionary origin (Mruk and Kobayashi, 2014). Type II TAs, which are constituted by two protein components, are the best characterized (Chan et al., 2016; Kędzierska and Hayes, 2016). Activation of TAs in response to stress was thought to be a way to either eliminate part of the population in benefit of the rest [i.e., “altruistic” cell death (Engelberg-Kulka and Glaser, 1999)], or to reduce the metabolic load during adverse conditions by slowing or arresting cell growth (Maisonneuve and Gerdes, 2014). Chromosomal TAs have also been associated with several bacterial processes, like biofilm formation, survival during infection of eukaryotic cells, defense against invading bacteriophages and entrance and exit into persistence (Goeders and Van Melder, 2014; Kędzierska and Hayes, 2016; Lobato-Márquez et al., 2016a). Since persistence is believed to be a major factor contributing to the chronic state of infections and tolerance to antibiotic treatments (Michiels et al., 2016), it was proposed that one of the roles of TAs was to contribute to dormancy, i.e., making the cells metabolically inactive (Pedersen et al., 2002; Christensen-Dalsgaard et al., 2010), a state that would lead to persistence due to triggering of the TAs (Maisonneuve et al., 2011; Maisonneuve and Gerdes, 2014). This attractive hypothesis was later considered as too simplistic (Ramisetty et al., 2016; Van Melder and Wood, 2017). Since the initial discovery of TA systems as plasmid maintenance systems and their subsequent identification in the chromosome, many examples underline the functional confluence of these systems irrespective of their location. In the following sections we will try to discuss this confluence based on a few well-characterized examples of type II TA systems (see **Table 1** for a summary of the TA systems covered in this review).

## PLASMID AND CHROMOSOMAL TOXIN–ANTITOXINS AS MEDIATORS OF POST-SEGREGATIONAL KILLING

Several observations indicate the existence of functional overlaps of TAs placed on plasmids or on chromosomes. A role of the chromosomal TAs in stabilization of integrated MGE or adjacent chromosomal regions has been demonstrated (Wozniak and Waldor, 2009), a role that is similar to their function in maintaining plasmid stability through PSK (Hayes, 2003). A novel TA pair designated *mosAT*, was shown to be responsible for maintaining the integrity of the ~100 kb SXT integrative and conjugative element (ICE) that confers resistance to multiple antibiotics in clinical isolates of *Vibrio cholerae* (Wozniak and Waldor, 2009). For a large MGE that

can integrate, excise, and transfer to other bacteria, the SXT ICE is remarkably stable, with loss of ICE estimated at only 1 in  $10^7$  cells (Wozniak and Waldor, 2009). The *mosAT* system has low basal transcriptional levels when SXT is integrated; however its expression is derepressed when SXT is in an extrachromosomal state and vulnerable to loss. Interestingly, a homolog of *mosAT* was located on an octopine-type Ti plasmid of *Agrobacterium tumefaciens* suggesting that this TA system may function to maintain the stability of plasmids as well as ICEs (Wozniak and Waldor, 2009). Another TA system designated *sgiTA* was recently shown to promote the maintenance of a multidrug resistant integrative and mobilizable *Salmonella* Genomic Island 1 (SGI1) in *Salmonella enterica* serovar Typhimurium (Huguet et al., 2016) in a manner similar to *mosAT* for SXT. Intriguingly, SGI1 is only transmissible in the presence of conjugative plasmids of the IncA/C group but paradoxically, SGI1 displayed incompatibility with the IncA/C plasmids. The *sgiTA* locus was shown to play an essential role in SGI1 stability particularly in the concomitant presence of a conjugative IncA/C plasmid when SGI1 is in an extrachromosomal state and is more likely to be lost (Huguet et al., 2016).

Further, elimination of cells that lose a chromosome has been demonstrated in the case of *V. cholerae* (Yuan et al., 2011). Like all vibrios, the *V. cholerae* genome consists of two chromosomes (Heidelberg et al., 2000), and the smaller chromosome II (ChrII, 1.07 Mbp) hosts a large 126 kb superintegron (SI) that gathers 100s of diverse gene cassettes, including antibiotic resistance genes, and contains 17 TA systems (Iqbal et al., 2015). Each of these cassettes is associated with a target recombination sequence (the *attC* site) and in a SI, 100s of gene cassettes with their *attC* sites are arranged in direct orientation, alluding to a likely inherent instability in these SIs. Nevertheless, SIs are remarkably stable, and in an earlier paper, it was elegantly demonstrated that two TA loci from the *Vibrio vulnificus* SI (*relBE1* and *parDE1*) stabilize the SI and prevent large scale deletions from occurring when the SI was devoid of TA loci (Szekeres et al., 2007). Chromosome-specific mechanisms exist to ensure the proper segregation of the two *V. cholerae* chromosomes in daughter cells. In the *V. cholerae* ChrII, the *parAB2* locus was essential for the partitioning of ChrII, and in a *parAB2* deletion mutant, ChrII was mislocalized leading to a complete loss of the entire ChrII in a fraction of the population (Yamaichi et al., 2007). Cells that lost ChrII were non-viable and underwent characteristic cytological changes including cell enlargement, nucleoid condensation and degradation (Yamaichi et al., 2007). It was subsequently shown that the three ParE toxins encoded by their respective *parDE* TA loci in the SI of ChrII were responsible for the PSK of cells that lost ChrII, closely mimicking PSK mediated by plasmid-encoded homologs (Yuan et al., 2011). A recent paper showed that all 17 TA loci in the *V. cholerae* SI were functional, expressed from their own native promoters, and were very specific – i.e., there was no cross-interaction between non-cognate toxins and antitoxins (Iqbal et al., 2015). These advocate for a major role of these 17 TA loci in the stabilization of the *V. cholerae* SI and to prevent the emergence of cells that lack ChrII; in other words, much like their plasmid-encoded homologs, the *V. cholerae* chromosomal



**TABLE 1** | Summary of the location and functions of toxin–antitoxin (TA) systems covered in this review.

TA	Host	Location of TA	Function	Reference
MosAT	<i>Vibrio cholerae</i>	Chromosomal; integrative and conjugative element (ICE)	Stability of ICE	Wozniak and Waldor, 2009
SgITA	<i>Salmonella enterica</i> serovar Typhimurium	Chromosomal; genomic island 1	Stability of genomic island	Huguet et al., 2016
RelBE1 ParDE1	<i>Vibrio vulnificus</i>	Chromosomal; superintegron	Superintegron stability	Szekeres et al., 2007
ParDE-1	<i>Vibrio cholerae</i>	Chromosome II, superintegron	Stability of superintegron as well as chromosome II	Iqbal et al., 2015
ParDE-2				
ParDE-3				
RelBE-1				
RelBE-2				
RelBE-3				
RelBE-4				
HigBA-1				
HigBA-2				
Phd-Doc				
0318-0319				
0332-0333				
0422-0423				
0477-0478				
0481-0482				
0486-0487				
0488-9489				
Epsilon-Zeta (ε-ζ)	<i>Streptococcus pyogenes</i>	Plasmid pSM19035	Plasmid stability	Camacho et al., 2002
PezAT	<i>Streptococcus pneumoniae</i>	Chromosomal, pathogenicity island 1	Stability of pathogenicity island, virulence factor	Chan et al., 2014; Iannelli et al., 2014
SezAT	<i>Streptococcus suis</i>	Chromosomal; pathogenicity island	Stability of pathogenicity island	Yao et al., 2015
AvrRxo1-Arc1	<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i>	Unknown	Virulence factor (type III secreted effector)	Triplett et al., 2016
CcdAB <sub>F</sub>	<i>Escherichia coli</i>	Plasmid F	Plasmid stability, persistence	Ogura and Hiraga, 1983; Tripathi et al., 2012
CcdAB <sub>O157</sub>	<i>Escherichia coli</i> O157:H7	Chromosomal	Persistence	Gupta et al., 2017
VapBC <sub>2ST</sub>	<i>Salmonella enterica</i> serovar Typhimurium	Plasmid pLST	Plasmid stability, virulence Plasmid stability	Lobato-Márquez et al., 2016b
CcdAB <sub>ST</sub>				

TA loci are also agents of PSK (Yuan et al., 2011; Iqbal et al., 2015).

Conversely, it has been shown that plasmid-encoded TAs can contribute to overcome stress, induce persistence, and could increase survival of bacterial cells during infection (Helaine et al., 2014), functions that were initially attributed to chromosomal TAs (Lobato-Márquez et al., 2016a). PSK mediated by TAs following the loss of genetic information associated to a plasmid can be considered as a situation of stress, to which the cells react by toxin activation. Transient activation of the *E. coli* F-plasmid-encoded CcdB toxin enhance the generation of drug-tolerant persister cells, and this process was found to be dependent on Lon protease and RecA (Tripathi et al., 2012). The F-plasmid-encoded *ccdAB<sub>F</sub>* locus has been well-established as a plasmid maintenance system (Jaffe et al., 1985) and the finding that it plays a role in persistence expands its function

as a transmissible persistence factor (Tripathi et al., 2012) (see below).

## PezAT AND ITS POTENTIAL ROLE IN VIRULENCE

Two recent articles published in *Frontiers* (Chan and Espinosa, 2016; Lobato-Márquez et al., 2016b) underline the concept that phenotypes associated to plasmid- or to chromosomally encoded TAs do overlap because independent of their location, toxins target similar functions and the TA operons are regulated and induced by similar conditions. The first example is provided by the pneumococcal *pezAT* operon (Khoo et al., 2007; Chan and Espinosa, 2016). The two genes constituting it are placed in the putative mobilizable pathogenicity island 1 (pneumococcal

pathogenicity island 1, PPI1) and that is found in nearly half of capsulated (virulent) *Streptococcus pneumoniae* strains (Chan et al., 2012). In some strains there is a second copy of the operon, located on the putative ICE Tn5253. A close homolog of the *PezAT* pair is the Epsilon-Zeta TA, which was discovered in the broad host-range plasmid pSM19035 of *Streptococcus pyogenes* (Camacho et al., 2002). Epsilon-Zeta differs from *PezAT* as the Epsilon antitoxin does not perform the transcriptional regulation of the operon, but rather by a third component, Omega, which also regulates the transcription of other genes encoded by the pSM19025 plasmid. In the *pezAT* operon, the *PezA* antitoxin performs the autoregulatory role (Khoo et al., 2007). Epsilon-Zeta plays an essential role in maintaining the stability of plasmid pSM19035 via PSK.

The Zeta/*PezT* toxins target the cell wall synthesis machinery by phosphorylating the peptidoglycan precursor, UDP-*N*-acetylglucosamine (UNAG) at the 3'-OH group of the *N*-acetylglucosamine moiety. The phosphorylated product, UDP-*N*-acetylglucosamine-3-phosphate (UNAG-3P), accumulates in the cytosol and inhibits MurA, which is the essential enzyme that catalyzes the initial step in peptidoglycan synthesis (Mutschler and Meinhart, 2011; Mutschler et al., 2011). Nevertheless, it was proposed that reduction in the UNAG levels is just one of several responses that is triggered by Zeta/*PezT* expression in response to stress (Tabone et al., 2014). The *pezAT* operon may play a role in stabilizing the MGE within the pneumococcal host (Chan et al., 2014; Iannelli et al., 2014). Further, a close homolog of *pezAT* was discovered in *Streptococcus suis* (designated *sezAT*), which was shown to be important for the stable inheritance of the Pathogenicity Island 1 (SsPI-1) (Yao et al., 2015). Pneumococcal strains that carry *pezAT* exhibit increased virulence; further, deletion of the operon led to pneumococcal cells exhibiting increased resistance to  $\beta$ -lactam antibiotics and to increased ability to take up homologous DNA by enhancing genetic competence (Chan and Espinosa, 2016). How *PezT* functions to increase pneumococcal virulence is currently unknown but it was postulated that activation of *PezT* during environmental stresses or the course of infection would result in inhibition of cell wall synthesis and subsequent lysis of a subpopulation of pneumococcal cells (Mutschler and Meinhart, 2011, 2013). The lysis of these cells would lead to the release of cellular components detrimental to the infected host such as pneumolysin. Interestingly, recent papers showed that a *PezT*/Zeta homolog, designated *AvrRxo1* from the plant pathogen *Xanthomonas oryzae* pv. *oryzicola* functions as a type III-secreted virulence factor which is toxic in plants and is bacteriostatic when expressed in *E. coli* (Han et al., 2015; Triplett et al., 2016). An *AvrRxo1* homolog from myxobacterium was also shown to trigger rapid cell death response in tobacco (Triplett et al., 2016). Intriguingly, although the *AvrRxo1* toxin was found to be a nucleotide kinase, its target is not UNAG like *PezT*/Zeta but rather, the coenzyme nicotinamide adenine dinucleotide (NAD) and its biochemical precursor, nicotinic acid adenine dinucleotide (NAAD) leading to the formation of unusual 3'-phosphorylated products, 3'-NADP and 3'-nicotinic acid adenine dinucleotide phosphate (3'-NADDP) (Schuebel et al., 2016). A recent paper showed

that 3'-NADP accumulates upon expression of *AvrRxo1* in tobacco and rice leaves infected with *AvrRxo1*-expressing strains of *Xanthomonas oryzae*, thus indicating that the *AvrRxo1* effector/toxin targets the coenzyme and redox carrier essential for central metabolic function of the host. However, the actual mechanism of 3'-NADP accumulation *in planta* is currently unknown as NAD and the conventional cofactor, 2'-NADP, are needed in 100s of essential reactions in the cell (Shidore et al., 2017). Hence, it could be possible that *PezT*/Zeta not only help in triggering the lysis of pneumococcal cells, the toxin itself may also be detrimental to the infected host cells. Indeed, it was recently shown that expression of the pneumococcal *pezT* toxin in the eukaryotic microalgae *Chlorella vulgaris* is lethal, leading to cellular damage and lysis (Ng et al., 2016). We await experimental results that would indicate if expression of *PezT*/Zeta in mammalian cells would be equally detrimental.

## FUNCTIONAL OVERLAPS BETWEEN CHROMOSOMAL AND PLASMID-ENCODED TA SYSTEMS

The second example is related to the role of TAs of *Salmonella enterica* serovar Typhimurium carrying the virulence plasmid pLST during bacterial infection (Lobato-Márquez et al., 2016b). One of the two TAs encoded by plasmid pLST is *vapBC<sub>ST</sub>*. This particular TA contributes to the successful colonization of recipient cells during infection, in conjunction with other type I and type II TAs encoded by the *Salmonella* chromosome (Lobato-Márquez et al., 2015). In addition to its role during infection, the plasmidic copy of *vapBC<sub>ST</sub>* contributes to the maintenance of the plasmid (Lobato-Márquez et al., 2016b). Interestingly, the chromosomal copy of this particular TA seems to be inactive, so that the role in infection was taken up by the plasmid-encoded copy. Curiously enough, the *VapC* toxin of *vapBC<sub>2ST</sub>* is active as a toxin, indicating that stabilization of pLST could be due to PSK (Lobato-Márquez et al., 2015).

An interesting example showing that location is compatible with different functions is provided by the first type II locus described, the *ccdAB<sub>F</sub>* operon encoded by plasmid F (Jaffe et al., 1985). This operon was reported to contribute to plasmid maintenance by killing plasmid free-segregants; PSK was the result of Lon protease-mediated degradation of the *CcdA* antitoxin and the subsequent activation of the anti-topoisomerase activity of the toxin *CcdB* (Ogura and Hiraga, 1983). In addition to its role in plasmid maintenance, *ccdAB<sub>F</sub>* was shown to contribute to bacterial persistence (Tripathi et al., 2012), a role that was also proposed for several chromosomal TA systems (Maisonneuve et al., 2011) and that has been questioned as reductionist recently (Ramisetty et al., 2016; Van Melderen and Wood, 2017). Furthermore, the *ccdAB<sub>ST</sub>* system of plasmid pLST seems to participate in plasmid maintenance beyond PSK because, in spite of carrying a single point mutation that inactivates the anti-topoisomerase activity of *CcdB<sub>ST</sub>*, it contributes significantly to the stabilization of the virulence plasmid by a yet to be identified mechanism

(Lobato-Márquez et al., 2016b). Functional interactions between co-existing *ccd* systems in plasmid and chromosome reported to be present in the pathogenic *E. coli* strain O157:H7 added further versatility to this system (Wilbaux et al., 2007). The chromosomally encoded *ccdAB* genes, like the plasmidic ones, has a toxin that target DNA gyrase and an antitoxin that is degraded by the Lon protease; however, both TAs seem to have evolved to achieve different functions: only the plasmidic antitoxin is able to neutralize the chromosomal toxin but not *vice versa*, and only the plasmidic TA is able to promote plasmid maintenance by PSK (Wilbaux et al., 2007). Nevertheless, a recent paper has shown that the chromosomal *ccdAB*<sub>O157</sub> system from *E. coli* O157:H7 also function in the formation of persister cells much like its F-plasmid-encoded counterpart even though the CcdB<sub>O157</sub> toxin displayed lower toxicity and has fivefold lower affinity for DNA gyrase compared to CcdB<sub>F</sub> (Gupta et al., 2017).

Plasmid maintenance linked to the coordination of TAs with plasmid replication was initially reported by the Díaz-Orejas laboratory on the *kis-kid* TA encoded by plasmid R1 (Ruiz-Echevarría et al., 1995a,b), and further analyzed (Pimentel et al., 2005; López-Villarejo et al., 2012, 2015). Additional work revealed a further coordination of the *kis-kid* TA with cell cycle functions (Pimentel et al., 2014). On the whole the above work supports that, failures in plasmid R1 replication reduces the levels of the Kis antitoxin and increases the activity of the Kid toxin. This results in: (i) the rescue of plasmid replication mediated by Kid-dependent decrease in the expression levels of CopB, a secondary inhibitor of plasmid replication, and (ii) the decrease in the levels of key cell division proteins that allows the rescue of plasmid replication before cell division can occur.

Targeting by RNase toxins of mRNAs, tRNAs and rRNAs impact and remodel protein synthesis and are key to the stress response mediated by TAs (reviewed by Moll and Engelberg-Kulka, 2012; Cruz and Woychik, 2015). Targeting tRNA and remodeling of protein synthesis profile seem to be a general mechanism of stress response. Indeed, a recent publication (Chionh et al., 2016) reveals a mechanism related to response to oxidative stress and induction of persistence in *Mycobacterium bovis*. The mechanism implies modification of the tRNA anticodons for threonine or leucine. Due to these modifications, the tRNA will enable the efficient translation of particular proteins related to the oxidative stress response. These will lead, in turn, to: (i) remodeling the protein synthesis potential of the cell to respond to oxidative stress; (ii) preferential synthesis of a set of stress response proteins, and (iii) induction of persistence.

Most interestingly these stress response mechanisms seem to be universal and are shared by prokaryotes and eukaryotes. Induction of persistence by TAs in response to stress also involves reduction of the protein synthesis potential and selective synthesis of proteins required to achieve survival to the stress-inducing agent (reviewed by Moll and Engelberg-Kulka, 2012).

## CONCLUDING REMARKS

TAs were proposed to be part of the accessory genome but as more and more details of their biological function are uncovered, the importance of these (apparently) expendable genetic entities to the lifestyle of their hosts are becoming clearer. As we have shown in this review, the biological functions of TAs do overlap irrespective of their location in their host genome – i.e., whether they are chromosomally encoded or plasmid-borne. Initially implicated in maintaining the stability of plasmids via PSK, TAs have since been shown to mediate the stability of genomic islands and even chromosome II of *V. cholerae* by PSK. Both plasmid- and chromosomally encoded TAs have also been implicated in persistence and virulence of several pathogens. It is thus clear that these hitherto “expendable” genetic loci have successfully integrated into their hosts’ cellular regulatory network, enabling their hosts to better adapt to their distinctive environmental niches.

## AUTHOR CONTRIBUTIONS

RD-O, ME, and CCY designed the outline of the review. RD-O wrote the first draft, and all authors worked on it until the production of the final version.

## FUNDING

While this review was being written, the authors participated in projects funded by MINECO-BIO2015-69085-REDC (to RD-O and ME) and FRGS/1/2016/SKK11/UNISZA/01/1 (to CCY).

## ACKNOWLEDGMENT

Thanks are due to Damián Lobato-Márquez for his critical reading of the manuscript.

## REFERENCES

- Bravo, A., Ortega, S., De Torriontegui, G., and Díaz, R. (1988). Killing of *Escherichia coli* cells modulated by components of the stability system ParD of plasmid R1. *Mol. Gen. Genet.* 215, 146–151. doi: 10.1007/BF00331316
- Camacho, A. G., Misselwitz, R., Behlke, J., Ayora, S., Welfle, K., Meinhardt, A., et al. (2002). *In vitro* and *in vivo* stability of the *epsilon*<sub>2</sub> protein complex of the broad host-range *Streptococcus pyogenes* pSM19035 addiction system. *Biol. Chem.* 383, 1701–1713. doi: 10.1515/BC.2002.191
- Chan, W. T., and Espinosa, M. (2016). The *Streptococcus pneumoniae* *pezAT* toxin-antitoxin system reduces  $\beta$ -lactam resistance and genetic competence. *Front. Microbiol.* 7:1322. doi: 10.3389/fmicb.2016.01322
- Chan, W. T., Espinosa, M., and Yeo, C. C. (2016). Keeping the wolves at bay: antitoxins of prokaryotic type II toxin-antitoxin systems. *Front. Mol. Biosci.* 3:9. doi: 10.3389/fmolb.2016.00009
- Chan, W. T., Moreno-Córdoba, I., Yeo, C. C., and Espinosa, M. (2012). Toxin-antitoxin genes of the gram-positive pathogen *Streptococcus pneumoniae*: so few and yet so many. *Microbiol. Mol. Biol. Rev.* 76, 773–791. doi: 10.1128/MMBR.00030-12
- Chan, W. T., Yeo, C. C., Sadowy, E., and Espinosa, M. (2014). Functional validation of putative toxin-antitoxin genes from the Gram-positive pathogen *Streptococcus pneumoniae*: *phd-doc* is the fourth bona-fide operon. *Front. Microbiol.* 5:677. doi: 10.3389/fmicb.2014.00677

- Chionh, Y. H., Mcbee, M., Babu, I. R., Hia, F., Lin, W., Zhao, W., et al. (2016). tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat. Commun.* 7:13302. doi: 10.1038/ncomms13302
- Christensen, S. K., and Gerdes, K. (2003). RelE toxins from bacteria and archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol. Microbiol.* 48, 1389–1400. doi: 10.1046/j.1365-2958.2003.03512.x
- Christensen-Dalsgaard, M., Jørgensen, M. G., and Gerdes, K. (2010). Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol. Microbiol.* 75, 333–348. doi: 10.1111/j.1365-2958.2009.06969.x
- Cooper, T. F., and Heinemann, J. A. (2000). Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12643–12648. doi: 10.1073/pnas.220077897
- Cruz, J. W., and Woychik, N. A. (2015). tRNAs taking charge. *Pathog. Dis.* 74:ftv117. doi: 10.1093/femspd/ftv117
- Engelberg-Kulka, H., and Glaser, G. (1999). Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.* 53, 43–70. doi: 10.1146/annurev.micro.53.1.43
- Gerdes, K., Bech, F. W., Jørgensen, S. T., Løbner-Olesen, A., Rasmussen, P. B., Atlung, T., et al. (1986a). Mechanism of postsegregational killing by the *hok* gene product of the *parB* system of plasmid R1 and its homology with the *relF* gene product of the *Escherichia coli relB* operon. *EMBO J.* 5, 2023–2029.
- Gerdes, K., Rasmussen, P. B., and Molin, S. (1986b). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3116–3120.
- Goeders, N., and Van Melderen, L. (2014). Toxin-antitoxin systems as multilevel interaction systems. *Toxins* 6, 304–324. doi: 10.3390/toxins6010304
- Gupta, K., Tripathi, A., Sahu, A., and Varadarajan, R. (2017). Contribution of the chromosomal *ccdAB* operon to bacterial drug tolerance. *J. Bacteriol.* doi: 10.1128/jb.00397-17 [Epub ahead of print].
- Han, Q., Zhou, C., Wu, S., Liu, Y., Triplett, L., Miao, J., et al. (2015). Crystal structure of *Xanthomonas* AvrXo1-ORF1, a Type III effector with a polynucleotide kinase domain, and its interactor AvrXo1-ORF2. *Structure* 23, 1900–1909. doi: 10.1016/j.str.2015.06.030
- Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 301, 1496–1499. doi: 10.1126/science.1088157
- Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483. doi: 10.1038/35020000
- Helaine, S., Cheverton, A. M., Watson, K. G., Faure, L. M., Matthews, S. A., and Holden, D. W. (2014). Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. *Science* 343, 204–208. doi: 10.1126/science.1244705
- Huguet, K. T., Gonnet, M., Doublet, B., and Cloeckaert, A. (2016). A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable *Salmonella* Genomic Island 1. *Sci. Rep.* 6:32285. doi: 10.1038/srep32285
- Iannelli, F., Santoro, F., Oggioni, M. R., and Pozzi, G. (2014). Nucleotide sequence analysis of integrative conjugative element *Tn5253* of *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 58, 1235–1239. doi: 10.1128/AAC.01764-13
- Iqbal, N., Guérout, A.-M., Krin, E., Le Roux, F., and Mazel, D. (2015). Comprehensive functional analysis of the 18 *Vibrio cholerae* N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. *J. Bacteriol.* 197, 2150–2159. doi: 10.1128/jb.00108-15
- Jaffe, A., Oruga, T., and Hiraga, S. (1985). Effects of the *ccd* function of the F plasmid on bacterial growth. *J. Bacteriol.* 163, 841–849.
- Kędzierska, B., and Hayes, F. (2016). Emerging roles of Toxin-Antitoxin modules in bacterial pathogenesis. *Molecules* 21:E790. doi: 10.3390/molecules21060790
- Khoo, S. K., Loll, B., Chan, W. T., Shoeman, R. L., Ngoo, L., Yeo, C. C., et al. (2007). Molecular and structural characterization of the *PezAT* chromosomal toxin-antitoxin system of the human pathogen *Streptococcus pneumoniae*. *J. Biol. Chem.* 282, 19606–19618. doi: 10.1074/jbc.M701703200
- Lobato-Márquez, D., Díaz-Orejas, R., and García-Del Portillo, F. (2016a). Toxin-antitoxins and bacterial virulence. *FEMS Microbiol. Rev.* 40, 592–609. doi: 10.1093/femsre/fuw022
- Lobato-Márquez, D., Molina-García, L., Moreno-Córdoba, I., García-Del Portillo, F., and Díaz-Orejas, R. (2016b). Stabilization of the virulence plasmid pSLT of *Salmonella* Typhimurium by three maintenance systems and its evaluation by using a new stability test. *Front. Mol. Biosci.* 3:66. doi: 10.3389/fmolb.2016.00066
- Lobato-Márquez, D., Moreno-Córdoba, I., Figueroa, V., Díaz-Orejas, R., and García-Del Portillo, F. (2015). Distinct type I and type II toxin-antitoxin modules control *Salmonella* lifestyle inside eukaryotic cells. *Sci. Rep.* 5:9374. doi: 10.1038/srep09374
- López-Villarejo, J., Diago-Navarro, E. E., Hernández-Arriaga, A. M., and Díaz-Orejas, R. (2012). Kis antitoxin couples plasmid R1 replication and *parD* (*kis*, *kid*) maintenance modules. *Plasmid* 67, 118–127. doi: 10.1016/j.plasmid.2011.12.015
- López-Villarejo, J., Lobato-Márquez, D., and Díaz-Orejas, R. (2015). Coupling between the basic replicon and the Kis-Kid maintenance system of plasmid R1: modulation by Kis antitoxin levels and involvement in control of plasmid replication. *Toxins* 7, 478–492. doi: 10.3390/toxins7020478
- Maisonneuve, E., and Gerdes, K. (2014). Molecular mechanisms underlying bacterial persisters. *Cell* 157, 539–548. doi: 10.1016/j.cell.2014.02.050
- Maisonneuve, E., Shakespeare, L. J., Jørgensen, M. G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13206–13211. doi: 10.1073/pnas.1100186108
- Michiels, J. E., Van Den Bergh, B., Verstraeten, N., and Michiels, J. (2016). Molecular mechanisms and clinical implications of bacterial persistence. *Drug Resist. Updat.* 29, 76–89. doi: 10.1016/j.drug.2016.10.002
- Moll, I., and Engelberg-Kulka, H. (2012). Selective translation during stress in *Escherichia coli*. *Trends Biochem. Sci.* 11, 493–498. doi: 10.1016/j.tibs.2012.07.007
- Mruk, I., and Kobayashi, I. (2014). To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.* 42, 70–86. doi: 10.1093/nar/gkt711
- Mutschler, H., Gebhardt, M., Shoeman, R. L., and Meinhardt, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol.* 9:e1001033. doi: 10.1371/journal.pbio.1001033
- Mutschler, H., and Meinhardt, A. (2011).  $\epsilon/\zeta$  systems: their role in resistance, virulence, and their potential for antibiotic development. *J. Mol. Med.* 89, 1183–1194. doi: 10.1007/s00109-011-0797-4
- Mutschler, H., and Meinhardt, A. (2013). “Type II toxin-antitoxin loci: the epsilon/zeta family,” in *Prokaryotic Toxin-Antitoxins*, ed. K. Gerdes (Berlin: Springer), 205–223.
- Ng, S. L., Hari Krishna, J. A., Abu Bakar, F., Yeo, C. C., and Cha, T. S. (2016). Heterologous expression of the *Streptococcus pneumoniae* *yoeB* and *pezT* toxin genes is lethal in *Chlorella vulgaris*. *Algal Res.* 19, 21–29. doi: 10.1016/j.algal.2016.07.011
- Ogura, T., and Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. U.S.A.* 80, 4784–4788. doi: 10.1073/pnas.80.15.4784
- Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* 12, 208–214. doi: 10.1038/nchembio.2044
- Pedersen, K., Christensen, K. S., and Gerdes, K. (2002). Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Mol. Microbiol.* 45, 501–510. doi: 10.1046/j.1365-2958.2002.03027.x
- Pimentel, B., Madine, M. A., and De La Cueva-Mendez, G. (2005). Kid cleaves specific mRNAs at UUACU sites to rescue the copy number of plasmid R1. *EMBO J.* 24, 3459–3469. doi: 10.1038/sj.emboj.7600815
- Pimentel, B., Nair, R., Bermejo-Rodríguez, C., Preston, M. A., Agu, C. A., Wang, X., et al. (2014). Toxin Kid uncouples DNA replication and cell division to enforce retention of plasmid R1 in *Escherichia coli* cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2734–2739. doi: 10.1073/pnas.1308241111
- Ramisetty, B. C. M., Ghosh, D., Roy Chowdhury, M., and Santhosh, R. S. (2016). What Is the link between stringent response, endoribonuclease encoding Type II Toxin-Antitoxin systems and persistence? *Front. Microbiol.* 7:1882. doi: 10.3389/fmicb.2016.01882
- Ruiz-Echevarría, M. J., De La Torre, M. A., and Díaz-Orejas, R. (1995a). A mutation that decreases the efficiency of plasmid R1 replication leads to the activation of *parD*, a killer stability system of the plasmid. *FEMS Microbiol. Lett.* 130, 129–135. doi: 10.1111/j.1574-6968.1995.tb07709.x
- Ruiz-Echevarría, M. J., Giménez-Gallego, G., Sabariego-Jareño, R., and Díaz-Orejas, R. (1995b). Kid, a small protein of the *parD* stability system of plasmid



- R1 is an inhibitor of DNA replication acting at the initiation of DNA synthesis. *J. Mol. Biol.* 247, 568–577. doi: 10.1016/S0022-2836(05)80138-X
- Schuebel, F., Rucker, A., Edelmann, D., Schessner, J., Brieke, C., and Meinhart, A. (2016). 3'-NADP and 3'-NAADP, two metabolites formed by the bacterial Type III effector AvrXo1. *J. Biol. Chem.* 291, 22868–22880. doi: 10.1074/jbc.M116.751297
- Shidore, T., Broeckling, C. D., Kirkwood, J. S., Long, J. J., Miao, J., Zhao, B., et al. (2017). The effector AvrXo1 phosphorylates NAD in planta. *PLOS Pathog.* 13:e1006442. doi: 10.1371/journal.ppat.1006442
- Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe-Magnus, D. A. (2007). Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol. Microbiol.* 63, 1588–1605. doi: 10.1111/j.1365-2958.2007.05613.x
- Tabone, M., Liou, V. S., Ayora, S., Machón, C., and Alonso, J. C. (2014). Role of Toxin  $\zeta$  and starvation responses in the sensitivity to antimicrobials. *PLoS ONE* 9:e86615. doi: 10.1371/journal.pone.0086615
- Tripathi, A., Dewan, P. C., Barua, B., and Varadarajan, R. (2012). Additional role for the ccd operon of F-plasmid as a transmissible persistence factor. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12497–12502. doi: 10.1073/pnas.1121217109
- Triplett, L. R., Shidore, T., Long, J., Miao, J., Wu, S., Han, Q., et al. (2016). AvrXo1 is a bifunctional Type III secreted effector and Toxin-Antitoxin system component with homologs in diverse environmental contexts. *PLoS ONE* 11:e0158856. doi: 10.1371/journal.pone.0158856
- Tsuchimoto, S., Ohtsubo, H., and Ohtsubo, E. (1988). Two genes, *pemK* and *pemI*, responsible for stable maintenance of resistance plasmid R100. *J. Bacteriol.* 170, 1461–1466. doi: 10.1128/jb.170.4.1461-1466.1988
- Van Melder, L., and Wood, T. K. (2017). Commentary: What is the link between stringent response, endoribonuclease encoding Type II Toxin-Antitoxin systems and persistence? *Front. Microbiol.* 8:191. doi: 10.3389/fmicb.2017.00191
- Wilbaux, M., Mine, N., Guerout, A.-M., Mazel, D., and Van Melder, L. (2007). Functional interactions between coexisting toxin-antitoxin systems of the *ccd* family in *Escherichia coli* O157:H7. *J. Bacteriol.* 189, 2712–2719. doi: 10.1128/jb.01679-06
- Wozniak, R. A. F., and Waldor, M. K. (2009). A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.* 5:e1000439. doi: 10.1371/journal.pgen.1000439
- Yamaichi, Y., Fogel, M. A., and Waldor, M. K. (2007). *par* genes and the pathology of chromosome loss in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 630–635. doi: 10.1073/pnas.0608341104
- Yao, X., Chen, T., Shen, X., Zhao, Y., Wang, M., Rao, X., et al. (2015). The chromosomal *SezAT* toxin-antitoxin system promotes the maintenance of the *SsPI-1* pathogenicity island in epidemic *Streptococcus suis*. *Mol. Microbiol.* 98, 243–257. doi: 10.1111/mmi.13116
- Yarmolinsky, M. (1995). Programmed cell death in bacterial populations. *Science* 267, 836–838. doi: 10.1126/science.7846528
- Yuan, J., Yamaichi, Y., and Waldor, M. K. (2011). The three *Vibrio cholerae* chromosome II-encoded *ParE* toxins degrade chromosome I following loss of chromosome II. *J. Bacteriol.* 193, 611–619. doi: 10.1128/jb.01185-10

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Díaz-Orejas, Espinosa and Yeo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



## OPEN ACCESS

## Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

## Reviewed by:

Christopher Morton Thomas,  
University of Birmingham,  
United Kingdom  
Elisabeth Grohmann,  
Beuth University of Applied Sciences,  
Germany  
Fabián Lorenzo,  
Universidad de La Laguna, Spain

## \*Correspondence:

Wilfried J. J. Meijer  
wmeijer@cbm.csic.es

## † Present address:

Gayetri Ramachandran,  
Synthetic Biology (G-5), Institute  
Pasteur, Paris, France

‡ These authors have contributed  
equally to this work.

## Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 03 August 2017

Accepted: 19 October 2017

Published: 03 November 2017

## Citation:

Miguel-Arribas A, Hao J-A,  
Luque-Ortega JR, Ramachandran G,  
Val-Calvo J, Gago-Córdoba C,  
González-Álvarez D, Abia D,  
Alfonso C, Wu LJ and Meijer WJJ  
(2017) The *Bacillus subtilis*  
Conjugative Plasmid pLS20 Encodes  
Two Ribbon-Helix-Helix Type Auxiliary  
Relaxosome Proteins That Are  
Essential for Conjugation.  
Front. Microbiol. 8:2138.  
doi: 10.3389/fmicb.2017.02138

# The *Bacillus subtilis* Conjugative Plasmid pLS20 Encodes Two Ribbon-Helix-Helix Type Auxiliary Relaxosome Proteins That Are Essential for Conjugation

Andrés Miguel-Arribas<sup>1‡</sup>, Jian-An Hao<sup>1,2‡</sup>, Juan R. Luque-Ortega<sup>3‡</sup>,  
Gayetri Ramachandran<sup>1†</sup>, Jorge Val-Calvo<sup>1</sup>, César Gago-Córdoba<sup>1</sup>,  
Daniel González-Álvarez<sup>1</sup>, David Abia<sup>1</sup>, Carlos Alfonso<sup>3</sup>, Ling J. Wu<sup>4</sup> and  
Wilfried J. J. Meijer<sup>1\*</sup>

<sup>1</sup> Department of Virology and Microbiology, Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Instituto de Biología Molecular “Eladio Viñuela” (CSIC), Autonomous University of Madrid, Madrid, Spain, <sup>2</sup> The Institute of Seawater Desalination and Multipurpose Utilization (SOA), Tianjin, China, <sup>3</sup> Centro de Investigaciones Biológicas (CSIC), Madrid, Spain, <sup>4</sup> Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle Upon Tyne, United Kingdom

Bacterial conjugation is the process by which a conjugative element (CE) is transferred horizontally from a donor to a recipient cell via a connecting pore. One of the first steps in the conjugation process is the formation of a nucleoprotein complex at the origin of transfer (*oriT*), where one of the components of the nucleoprotein complex, the relaxase, introduces a site- and strand specific nick to initiate the transfer of a single DNA strand into the recipient cell. In most cases, the nucleoprotein complex involves, besides the relaxase, one or more additional proteins, named auxiliary proteins, which are encoded by the CE and/or the host. The conjugative plasmid pLS20 replicates in the Gram-positive Firmicute bacterium *Bacillus subtilis*. We have recently identified the relaxase gene and the *oriT* of pLS20, which are separated by a region of almost 1 kb. Here we show that this region contains two auxiliary genes that we name *aux1*<sub>LS20</sub> and *aux2*<sub>LS20</sub>, and which we show are essential for conjugation. Both *Aux1*<sub>LS20</sub> and *Aux2*<sub>LS20</sub> are predicted to contain a Ribbon-Helix-Helix DNA binding motif near their N-terminus. Analyses of the purified proteins show that *Aux1*<sub>LS20</sub> and *Aux2*<sub>LS20</sub> form tetramers and hexamers in solution, respectively, and that they both bind preferentially to *oriT*<sub>LS20</sub>, although with different characteristics and specificities. *In silico* analyses revealed that genes encoding homologs of *Aux1*<sub>LS20</sub> and/or *Aux2*<sub>LS20</sub> are located upstream of almost 400 relaxase genes of the Rel<sub>LS20</sub> family (MOB<sub>L</sub>) of relaxases. Thus, *Aux1*<sub>LS20</sub> and *Aux2*<sub>LS20</sub> of pLS20 constitute the founding member of the first two families of auxiliary proteins described for CEs of Gram-positive origin.

**Keywords:** conjugation, relaxosome, auxiliary protein, DNA binding protein, Ribbon-Helix-Helix, antibiotic resistance, Firmicutes, horizontal gene transfer

## INTRODUCTION

Bacteria exchange genetic material at gross scale, even between distantly related species, via different routes collectively called horizontal gene transfer (HGT) (for review see, Ochman et al., 2000; Frost et al., 2005; Thomas and Nielsen, 2005; Boto, 2010). Horizontal exchange of DNA provides bacteria instantly with a new set of gene(s) and hence is an important driver for the rapid adaptation and evolution of bacteria. Among the genes that are spread by HGT are those responsible for antibiotic resistance (AR), which poses a serious and increasingly worrisome economic and health problem at a global scale. Three main mechanisms are responsible for HGT: transformation through natural competence, transduction via bacterial phage, and conjugation (Ochman et al., 2000; Frost et al., 2005; Thomas and Nielsen, 2005). Of these, conjugation appears to be the route that is predominantly responsible for spreading AR genes (Mazel and Davies, 1999; Waters, 1999; Norman et al., 2009; Davies and Davies, 2010). Conjugation is the process by which a conjugative element (CE) is transferred from a donor cell to a recipient cell through a dedicated transportation pore connecting both cells. CEs contain all the genes required for processing the DNA, establishing contact with the recipient cell, those encoding the structural proteins of the connecting pore as well as those for transporting the DNA. CEs can be integrated in a bacterial chromosome or be present on plasmids, which are named integrative and conjugative elements (ICEs) and conjugative plasmids, respectively. Due to the enormous numbers and density of microbes and the constant replenishment of bacteria upon the intake of food and liquids, the intestinal gut of humans and animals is a niche that is particularly apt for emerging, pooling, and spreading AR (Sommer et al., 2009, 2010; Forsberg et al., 2012; Penders et al., 2013).

Conjugative elements are commonly present in Gram-positive (G+) and Gram-negative (G-) bacteria and the basic concepts of the transfer process are conserved (Alvarez-Martinez and Christie, 2009; De la Cruz et al., 2010; Smillie et al., 2010; Goessweiner-Mohr et al., 2013). However, whereas in most systems conjugation involves the transfer of a single DNA strand (see below), DNA is transferred in its double-stranded form during conjugation in G+ mycelial *Streptomyces* bacteria (Goessweiner-Mohr et al., 2013; Thoma and Muth, 2016), which is not further considered here. Conjugation starts with a process named mating pair formation (Mpf) in which a donor cell recognizes and interacts with a suitable recipient cell. Probably, this triggers the signal for processing the DNA of the CE and subsequent transfer of one of its strands, named T-strand, into the recipient cell. The sophisticated, multi-component pore connecting the donor and the recipient cell is named transferosome, which is a type IV secretion system (T4SS). The enzyme responsible for initiating the generation of the T-strand is a relaxase, a phosphodiesterase, that cleaves the DNA in a strand- and site-specific manner at a specific position called the *nic* site, which is located within the origin of transfer region (*oriT*). Relaxase-mediated cleavage generates a hydroxyl group at 3' end of the *nic* site which functions as a primer for DNA elongation; i.e., the relaxase initiates a rolling-circle type of DNA

replication (also named DNA transfer replication [Dtr]). Upon nicking, the relaxase remains covalently attached to the 5'-end of the nicked T-strand which is then transferred, together with the attached T-strand, into the recipient cell. In most cases the active site residue that becomes covalently attached to the T-strand concerns a tyrosine. However, very recently it has been shown that relaxases of the MOB<sub>V</sub> family employ a histidine instead of a tyrosine residue to nick the DNA (Pluta et al., 2017). Due to its crucial role in conjugation, relaxases have attained considerable attention and several of them have been characterized in detail at the biochemical, functional and structural levels. In some cases, for instance ICEBs1 of *Bacillus subtilis* and the broad host range conjugative plasmid pIP501, the relaxase is the only protein that is required for processing the DNA (Kopec et al., 2005; Lee and Grossman, 2007; Grohmann et al., 2016). However, in the majority of cases additional protein(s), encoded either by the CE or the host, bind to the *oriT* and are involved in processing of the DNA. The nucleoprotein complex at *oriT* formed by the relaxase and additional proteins is called the relaxosome, and the additional proteins are named auxiliary or accessory proteins. Although their name may suggest that they play secondary role(s) in the processing reaction, most if not all of the auxiliary proteins studied so far have been shown to be essential for conjugation.

Most conjugation studies are based on CE present in G- bacteria, with knowledge on conjugation-related aspects in G+ bacteria lagging far behind. This is especially the case for auxiliary proteins (see Discussion). In our laboratory we study the conjugative plasmid pLS20 which was originally isolated from the Gram+ Firmicute bacterium *B. subtilis* natto IFO3335 (Tanaka et al., 1977). This strain is used for the fermentation of soybeans to produce "natto," a popular dish in South Asia, and hence it is conceivable that pLS20 or relatives play a role in the conjugation-mediated HGT in the gut of humans and animals. A derivative of pLS20 containing a chloramphenicol-resistance gene, pLS20cat, has been constructed (Itaya et al., 2006) and its sequence has been determined in our lab and in the lab of M. Itaya (Mitsuhiro Itaya, Keio University, Japan). All conjugation genes are located in one large operon spanning genes 28 till 74 according to our nomenclature (Singh et al., 2013). pLS20cat genes 25-27 are involved in regulating the expression of the conjugation genes (Singh et al., 2013; Ramachandran et al., 2014). Recently, we have identified and characterized the relaxase (gene 58) and the *oriT* of pLS20cat, which we named Rel<sub>LS20</sub> and *oriT*<sub>LS20</sub>, respectively (Ramachandran et al., 2017). Contrary to many other plasmids, the relaxase gene and *oriT* are located within its large conjugation operon, and Rel<sub>LS20</sub> turned out to be the founding member of a novel relaxase family containing >800 members.

Here, we addressed the question whether pLS20cat contains auxiliary relaxosome genes. We demonstrate that genes 56 and 57, located in between the relaxase gene *rel*<sub>LS20</sub> and *oriT*<sub>LS20</sub> are two auxiliary genes that are essential for conjugation and denominated them as *aux1*<sub>LS20</sub> and *aux2*<sub>LS20</sub>, respectively. Both gene products were purified and biochemical analyses showed that one of them formed tetramers and the other hexamers in solution. We also show that the proteins bind to distinct DNA motifs present in *oriT*<sub>LS20</sub>. *In silico* analyses revealed that a large fraction of the relaxase genes coding for the MOB<sub>I</sub>

family of relaxases are preceded by genes encoding homologs of Aux1<sub>LS20</sub> and/or Aux2<sub>LS20</sub>. The findings obtained for Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> are placed in perspective with other auxiliary proteins of CE present in G+ and G− organisms.

## MATERIALS AND METHODS

### Bacterial Strains, Plasmids, Media and Oligonucleotides

*Escherichia coli* and *B. subtilis* strains were grown in Luria-Bertani (LB) liquid medium or on 1.5% LB agar plates. When appropriate, media were supplemented with the following antibiotics: ampicillin (100 µg/ml), erythromycin (1 and 150 µg/ml in *B. subtilis* and *E. coli*, respectively), chloramphenicol (5 µg/ml), spectinomycin (100 µg/ml), and kanamycin (10 and 30 µg/ml in *B. subtilis* and *E. coli*, respectively). *B. subtilis* strains used were isogenic with *B. subtilis* strain 168 and are listed in Supplementary Table S1. Plasmids and oligonucleotides used are listed in Supplementary Tables S2, S3, respectively. All oligonucleotides were purchased from Isogen Life Science, Netherlands.

### Transformation

*Escherichia coli* cells were transformed using standard methods (Sambrook et al., 1989). Preparation of competent *B. subtilis* cells and transformation were carried as described before (Bron et al., 1989). Transformants were selected on LB agar plates with appropriate antibiotics. pLS20cat encodes a protein, Rok<sub>LS20</sub>, that inhibits the development of competence by repressing *comK*, the key transcriptional activator of competence genes (Singh et al., 2012). Therefore, to manipulate genes on pLS20cat we prepared competent cells of a pLS20cat-harboring strain that contains a chromosomal P<sub>xyl</sub>-*comK* fusion (PKS56) using a standard protocol (Singh et al., 2012).

### Construction of Plasmids and Strains

The correctness of sequences of all cloned PCR fragments was confirmed by sequence analysis. Amplification by PCR of pLS20cat regions was performed using as template total DNA isolated from pLS20cat harboring strain PKS11. Details regarding the construction of integration vectors based on plasmids pDR110 (*amyE* integration vector with IPTG-inducible P<sub>spank</sub> promoter) or pAXO1 (*lacA* integration vector with xylose-inducible P<sub>xyl</sub> promoter) are given in Supplementary Table S2. In summary, gene 56 was cloned under the control of the P<sub>xyl</sub> promoter or the P<sub>spank</sub> promoter. In addition, genes 56-57-58, genes 57-58, or gene 58 were cloned behind the P<sub>spank</sub> promoter. Plasmid DNA of the constructed pAXO1 and pDR110 derivatives was isolated from *E. coli* cells and then used to transform competent *B. subtilis* cells. Double-crossover integration into the chromosome was checked by PCR in the case of the pAXO1-derivatives. When pDR110 derivatives were used to transform competent *B. subtilis* cells, double cross over integration was tested by the loss of amylase activity. The pLS20cat genes 58 (*relLS20*), 57 (*aux1LS20*) and 56 (*aux2LS20*) were cloned in the *E. coli* expression vector pET28b+ to generate fusion genes

containing a C-terminal *his*<sub>(6)</sub> extension. Details regarding these cloning strategies are given in Supplementary Table S2. The resulting derivatives of pET28b+ were constructed using *E. coli* strain XL1-Blue. Once verified its correctness, the plasmids were transformed into *E. coli* strain BL21(DE3).

### Conjugation Assays

Conjugation was carried out in liquid medium as described previously (Singh et al., 2013). The effect of ectopic expression of a given gene placed under the control of the inducible P<sub>spank</sub> and/or P<sub>xyl</sub> promoter on conjugation was studied by adding the inducer (1 mM IPTG, 1% xylose) to prewarmed LB medium used to dilute overnight cultures of the donor cells.

### Analytical Ultracentrifugation Experiments

Sedimentation velocity (SV), sedimentation equilibrium (SE), and dynamic light scattering (DLS) assays and processing of the data, including estimations of molar masses of the relaxosome proteins from the hydrodynamic measurements, were carried out using the same conditions to those used before in the analysis of Rel<sub>LS20</sub> (Ramachandran et al., 2017).

### Over Expression and Purification of Recombinant Rel<sub>LS20</sub>, Aux1<sub>LS20</sub>, and Aux2<sub>LS20</sub> Containing a C-Terminal His<sub>(6)</sub> Tag

Recombinant versions of Rel<sub>LS20</sub>, Aux1<sub>LS20</sub>, and Aux2<sub>LS20</sub> were expressed and purified using similar protocols. In brief, *E. coli* BL21(DE3) cells containing plasmid pAND83 (*relLS20His*<sub>(6)</sub>), or pHJA56 (*aux1LS20His*<sub>(6)</sub>), or pHJA57 (*aux2LS20His*<sub>(6)</sub>) were inoculated in fresh LB media complemented with 30 µg/ml kanamycin and grown at 37°C with shaking (200 rpm). At an OD<sub>600</sub> of about 0.6, IPTG was added to a final concentration of 1 mM to induce the recombinant protein and growth was continued for 2 h. Next, cells were collected by centrifugation and processed as described before (Singh et al., 2012). The nickel-column purified proteins (>95% pure) were finally dialysed against buffer B (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM NaCl, 10 mM MgCl<sub>2</sub>, 7 mM β-mercaptoethanol, 50% v/v glycerol) and stored in aliquots at −80°C. Bradford assay and OD<sub>280</sub> determination were used to determine the protein concentrations.

### Gel Retardation Assays

Gel retardation assays were essentially carried out as described before (Singh et al., 2012). Thus, different DNA fragments were amplified by PCR using pLS20cat as template. The resulting PCR fragments were purified and 170 ng of DNA [200 or 362 bp] (with or without 220 ng of control DNA [176 bp]) were incubated on ice in binding buffer [20 mM Tris HCl pH 8, 1 mM EDTA, 5 mM MgCl<sub>2</sub>, 0.5 mM DTT, 100 mM KCl, 10% (v/v) glycerol, 0.05 mg ml<sup>−1</sup> BSA] without or with purified Aux1<sub>LS20</sub> or Aux2<sub>LS20</sub> to a fixed final concentration of 90 nM (Supplementary Figure S3) or using twofold increasing concentrations ranging from 0.09 to 5.76 µM (Figure 3) in a total volume of 16 µl. The



negative control, corresponding to bp numbers 63,774–63,950 of accession number NC\_015148.1, has an AT-content that is very similar to the AT content of the *oriT* fragment (61.4 vs. 61.1%). This DNA corresponds to sequences located inside a gene (gene 24), lowering the possibility that it harbors particular features for recruiting a transcriptional regulator or other DNA binding protein. In addition, it is predicted to lack a static bend. After careful mixing, samples were incubated for 20 min at 30°C, placed back on ice for 10 min, then loaded onto 2% agarose gel in 0.5XTBE. Electrophoresis was carried out in 0.5XTBE at 50 V at 4°C. Finally, the gel was stained with ethidium bromide, destained in 0.5XTBE and photographed with UV illumination.

## In Silico Analyses

### Identification of Mob<sub>L</sub> Members

Rel<sub>LS20</sub> was used as a query sequence to execute a psi-blast (version 2.6.1+) search against the NCBI nr protein database (July, 2017), allowing up to 10 rounds of reiteration with an e-value threshold of 1e-15 (Altschul et al., 1997, 2005; Schaffer et al., 2001) producing 1445 hits. The program “USEARCH” (version v10.0.240\_i86linux32) was then used to identify and remove redundant sequences showing 100% identity (Edgar, 2010), resulting in 1249 unique hits showing high similarity to Rel<sub>LS20</sub>.

### Identification of Putative Auxiliary Proteins

Protein sequences of Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> were used as query against the NCBI nr protein database (July 2017) using psi-blast (version 2.6.1+), with an e-value threshold of 1e-6 and 1e-7, respectively, until no new hits were retrieved. The sequence identifiers obtained from psi-blast, were crossed with the sequence identifiers preceding the MOB<sub>L</sub> family relaxase members, obtained from the nucleotide entries from they were translated.

### Prediction of Secondary Structure for Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> Homologs

Corresponding sequences were submitted to the RaptorX property web server (Wang et al., 2016) and predictions for  $\beta$ -strands and  $\alpha$ -helices along the sequences were plotted with “R”<sup>1</sup> (R Core Team, 2017).

## RESULTS

### Identification of Putative Relaxosome Genes of pLS20cat by *in Silico* Analysis

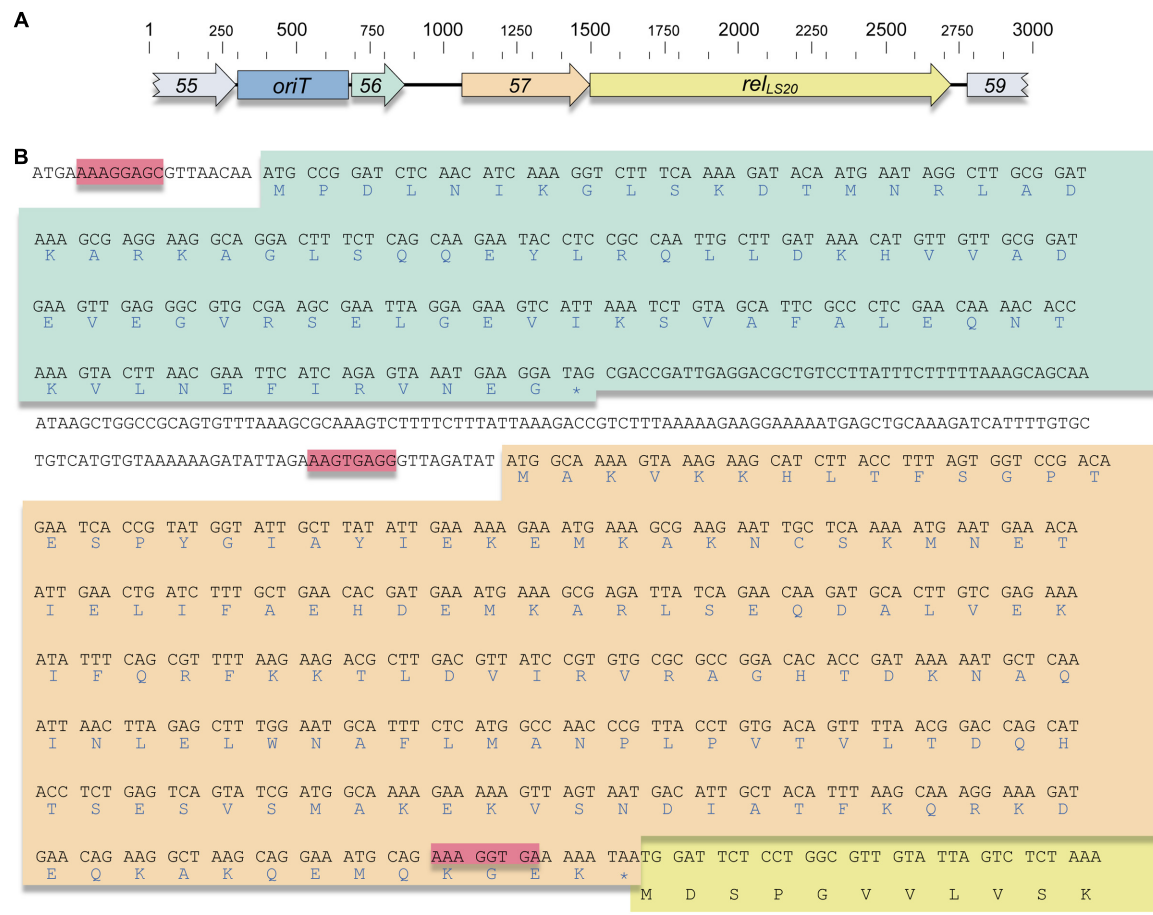
Recently, we have shown that pLS20cat gene 58 is essential for conjugation and that it encodes the relaxase, which we named Rel<sub>LS20</sub> (Ramachandran et al., 2017). In these studies we also identified the *nic* site of Rel<sub>LS20</sub> and delineated the functional *oriT*, named *oriT*<sub>LS20</sub>, to a region of 362 bp. Remarkably, *oriT*<sub>LS20</sub> and *rel*<sub>LS20</sub> are separated by a region of 865 bp, which has been annotated to contain two relatively small putative genes, designated genes 56 and 57 (Singh et al., 2013, see **Figure 1**

for a schematic view of this region). Often, but not always, conjugative plasmid-located relaxase genes are accompanied by small auxiliary relaxosome genes that generally are located upstream of the relaxase gene. This prompted us to investigate whether genes 56 and 57 might encode auxiliary relaxosome genes of pLS20cat. *In silico* analyses of pLS20cat genes 56 and 57 show that, firstly, *rel*<sub>LS20</sub> is translationally coupled to the preceding gene 57 [i.e., the stop (TAA) and start codon (ATG) of genes 57 and *rel*<sub>LS20</sub>, respectively, overlap; see **Figure 1**], and only a small intergenic region of 183 bp separates gene 57 from its preceding gene 56. Second, gene 56 and 57 are both small genes (79 and 147 codons, respectively). And third, the proteins encoded by these genes are both putative DNA binding proteins predicted to contain a Ribbon-Helix-Helix (RHH) motif in their N-terminal regions. An overview of the secondary structure prediction of both proteins and their homology with CopG, a paradigm of RHH DNA binding protein (Gomis-Ruth et al., 1998; Del Solar et al., 2002), is shown in Supplementary Figure S1. This figure shows that both Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> contain several lysine and arginine residues near the end of their predicted helix 1 and beginning of helix 2. The corresponding region in known RHH structures has been shown to be close to the phosphate backbone of the DNA (for example see, Schildbach et al., 1999). In summary, *in silico* analyses suggested that the two small genes 56 and 57 preceding the relaxase gene *rel*<sub>LS20</sub> may encode auxiliary relaxosome proteins.

### pLS20cat Genes 56 and 57 Are Essential for Conjugation

Previously, we engineered a derivative of pLS20cat, pLS20 $\Delta$ 56-58, in which the putative genes 56-57 together with the relaxase gene *rel*<sub>LS20</sub> (gene 58) have been deleted, and demonstrated that this plasmid was deficient in conjugation. Conjugation of pLS20 $\Delta$ 56-58 was restored when all three genes (56-58), were ectopically expressed from the IPTG-inducible P<sub>spank</sub> promoter at the chromosomal *amyE* locus, but not in the absence of gene 58, showing that Rel<sub>LS20</sub> was essential for conjugation (Ramachandran et al., 2017). We used a similar approach to study whether genes 56 and/or 57 were essential for conjugation. Thus, we constructed strain GR153, which harbors pLS20 $\Delta$ 56-58 and also contains *rel*<sub>LS20</sub> (gene 58), but not 56 and 57, under the control of the P<sub>spank</sub> promoter at the *amyE* locus. We then employed this strain as donor to determine the conjugation efficiencies using a standard protocol (see Materials and Methods). Strains PKS11, GR149 and GR150 were included as controls. As shown in **Table 1**, the efficiency of conjugation observed for the wild type plasmid pLS20cat was in the range of 10<sup>-3</sup>, which is similar to values reported previously under similar conditions (Singh et al., 2013; Ramachandran et al., 2014, 2017). As reported before (Ramachandran et al., 2017), conjugation was observed for pLS20 $\Delta$ 56-58 only when genes 56-58 were expressed from the chromosome (**Table 1**, strain GR149 and GR150). Importantly, no transconjugants were obtained when strain GR153 (*amyE*::P<sub>spank</sub>-*rel*<sub>LS20</sub>, pLS20 $\Delta$ 56-58) was used as donor in conjugation experiments, regardless of whether they were grown in the presence or absence of IPTG. These results

<sup>1</sup><https://www.R-project.org/>



**FIGURE 1 |** The pLS20cat relaxosome region. **(A)** Genetic organization of the pLS20cat genes 55–59. The genes, their sizes and orientation of transcription are indicated with arrows. Genes 55 and 59 are colored gray. Genes 56, 57, and 58 (*rel*<sub>LS20</sub>) are colored green, orange, and yellow, respectively. The same color code is used in “B,” as well as in **Figure 4** (see below). The 362 bp *oriT*<sub>LS20</sub> region is indicated with a blue box labeled *oriT*. Base pair numbering is given on the top. **(B)** DNA sequence of genes 56 and 57 and their deduced protein sequences. Stop codons are indicated with an asterisk and likely Ribosomal Binding sites (RBS) are highlighted with a red box. Note that genes 57 and *rel*<sub>LS20</sub> are translationally coupled. Only the first 11 codons of the *rel*<sub>LS20</sub> gene are given.

**TABLE 1 |** pLS20cat genes 56 and 57 are required for conjugation.

Strain	Genotype	Plasmid	Inductor	Conjugation efficiency*
PKS11	168	pLS20cat	–	7.9 × 10 <sup>−3</sup>
GR149	168	pLS20Δ56-58	–	<10 <sup>−7</sup>
GR150	168, <i>amyE</i> ::P <sub>spank</sub> 56-58	pLS20Δ56-58	–	<10 <sup>−7</sup>
			+	3.2 × 10 <sup>−3</sup>
GR153	168, <i>amyE</i> ::P <sub>spank</sub> 58	pLS20Δ56-58	–	<10 <sup>−7</sup>
			+	<10 <sup>−7</sup>
GR197	168, <i>amyE</i> ::P <sub>spank</sub> 57-58	pLS20Δ56-58	–	<10 <sup>−7</sup>
			+	<10 <sup>−7</sup>
GR200	168, <i>amyE</i> ::P <sub>spank</sub> 58, <i>lacA</i> ::P <sub>xyI</sub> -56	pLS20Δ56-58	–	<10 <sup>−7</sup>
			+	<10 <sup>−7</sup>
GR225	168, <i>amyE</i> ::P <sub>spank</sub> 57-58, <i>lacA</i> ::P <sub>xyI</sub> -56	pLS20Δ56-58	–	<10 <sup>−7</sup>
			+	5.8 × 10 <sup>−4</sup>

\*Conjugation efficiencies were calculated as transconjugants/donor, and correspond to the mean value of at least three independent experiments. When indicated, the inducer IPTG was added at a final concentration of 1 mM in the case of strains GR150 and GR197. In the case of GR200 and GR225, the final concentrations of the inducers was 1 mM (IPTG) and 1% (xylose).

showed that pLS20cat gene 56 and/or 57 are necessary for conjugation.

We next tested whether only one or both genes were required for conjugation. For this, we constructed the pLS20Δ56-58-harboring strains GR197 and GR200 in which *rel*<sub>LS20</sub> together with either gene 57 (strain GR197) or gene 56 (strain GR200) could be induced from the bacterial genome. When used as donor, no transconjugants were obtained for each strain regardless whether they were grown in the absence or presence of the inducer(s) (see **Table 1**), demonstrating that both genes are essential for conjugation.

In the above conjugation experiments, one or a combination of genes 56, 57, *rel*<sub>LS20</sub> was complemented by expressing them from the IPTG-inducible *P*<sub>spank</sub> promoter for all the strains except for strain GR200. In this strain *rel*<sub>LS20</sub> is controlled by *P*<sub>spank</sub> at the *amyE* locus and gene 56 by the xylose-inducible *P*<sub>xyl</sub> promoter at the *lacA* locus. To rule out the possibility that transconjugants were not obtained for donor strain GR200 because the genes were expressed from different promoters at a different locus, we constructed strain GR225 in which gene 56 was placed under the control of the *P*<sub>xyl</sub> promoter at *lacA*, and genes 57 and 58 under the control of the *P*<sub>spank</sub> promoter at *amyE*. Transconjugants were obtained for this strain when cells were grown in the presence of both inducers (**Table 1**), demonstrating that the gene products expressed from the two different promoters and chromosomal loci were all functional. These results demonstrate therefore that besides *rel*<sub>LS20</sub> genes 56 and 57 are also required for conjugation. Taking into account these results, together with the structural organization of these genes with respect to *rel*<sub>LS20</sub> and *oriT*<sub>LS20</sub>, the *in silico* analyses presented above, and additional evidence presented below, we conclude that pLS20cat gene 56 and 57 encode auxiliary relaxosome proteins which we name Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>, respectively.

## In Vitro Analysis of the Relaxosome Proteins Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>, and Rel<sub>LS20</sub>

### Oligomerization State Determined by Analytical Ultracentrifugation and DLS Techniques

To characterize the auxiliary relaxosome proteins *in vitro*, we purified Aux1<sub>LS20</sub> (Mw 10,601 Da) and Aux2<sub>LS20</sub> (Mw 18,605 Da) from *E. coli*, each fused to a His<sub>6</sub> tag at its C-terminus. We first determined the oligomerization state of the proteins, and also investigated putative interactions among them and with Rel<sub>LS20</sub>, using two complementary analytical ultracentrifugation approaches, i.e., SV and SE (**Figures 2A–D**), together with DLS experiments using the same experimental conditions.

Sedimentation profiles obtained by SV assays showed Aux1<sub>LS20</sub> as a single species with an experimental sedimentation coefficient of 2.5 S (*s*<sub>20,w</sub> = 2.9 S) compatible with a moderately elongated tetrameric form of the protein (*f*/*f*<sub>0</sub> = 1.5) (**Figure 2A**). Subsequent analysis of Aux1<sub>LS20</sub> gave a *D*-value of 52.5 ± 0.3 μm<sup>2</sup>/s. The obtained *S*- and *D*-values, once introduced in the Svedberg equation, yielded an apparent molar mass of 46,290 Da. SE data, best-fit analysis to single species model gave

an average molecular mass of 42,200 Da ± 300 Da, confirming that Aux1<sub>LS20</sub> is a tetramer in solution (**Figure 2B**).

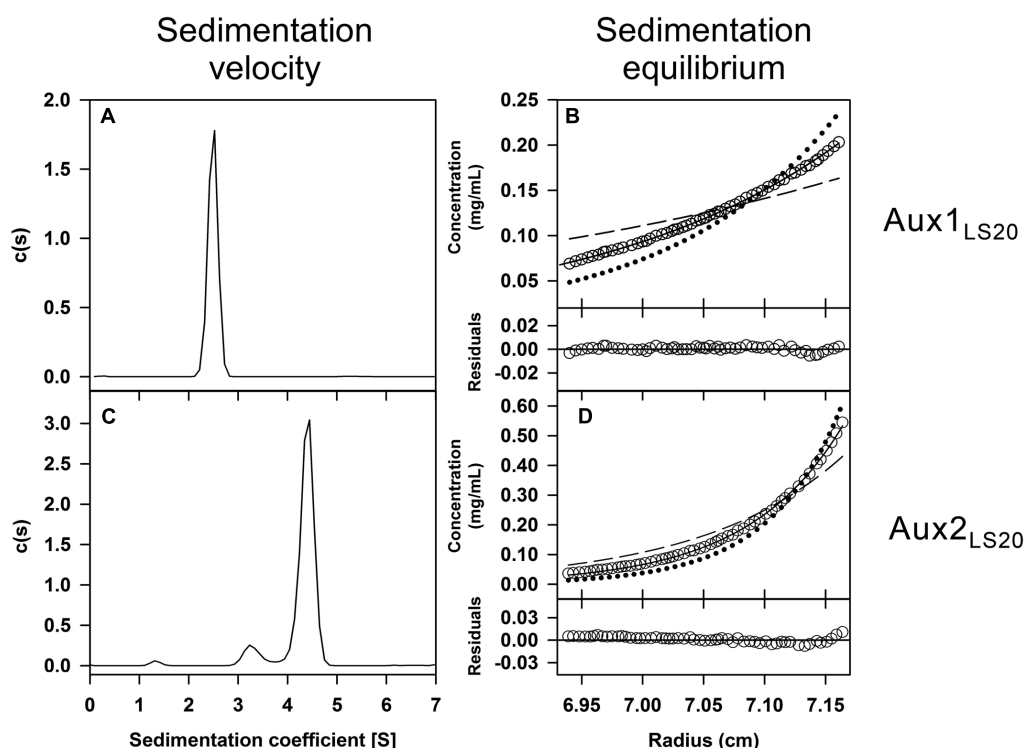
In the case of Aux2<sub>LS20</sub>, analysis of the sedimenting boundaries showed a sedimentation profile with a main peak corresponding to 90.0% of the total proteins at 4.4 S (*s*<sub>20,w</sub> = 5.1 S), together with a second peak at 3.3 S (*s*<sub>20,w</sub> = 3.8 S) encompassing 7% of the sample (**Figure 2C**). The *S*-value of the main peak is compatible with the theoretical behavior of a spherical Aux2<sub>LS20</sub> tetramer (*f*/*f*<sub>0</sub> = 1.2), as well as with a moderately elongated hexamer (*f*/*f*<sub>0</sub> = 1.6). DLS analysis of Aux2<sub>LS20</sub> yielded a *D* of 38.2 ± 1.0 μm<sup>2</sup>/s, which combined with the obtained *S*-value of 4.4 in the Svedberg formula resulted in an apparent molar mass of 113,400 Da that is very close to the molecular mass of Aux2<sub>LS20</sub> hexamers (111,630 Da). SE experiments were decisive for establishing the oligomerization state of Aux2<sub>LS20</sub>, as the best fit of the SE data gave an average molecular mass of 111,300 ± 1,200 Da, unequivocally demonstrating that Aux2<sub>LS20</sub> forms hexamers in solution (**Figure 2D**). In summary, the outcome of three complementary experimental approaches showed that Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> form tetramer and hexamers in solution, respectively.

Previously, we determined that purified Rel<sub>LS20</sub> forms monomers in solution (Ramachandran et al., 2017). To study possible interactions between the relaxosome proteins in solution we used combinations of Aux1<sub>LS20</sub>, Aux2<sub>LS20</sub> and Rel<sub>LS20</sub> and subjected these to SV experiments (Supplementary Figure S2). No additional peaks with increased *S*-values reflecting new protein hetero-complexes were obtained in any of the combinations tested implying that the relaxosome proteins of pLS20cat do not interact in solution, at least not under the conditions tested.

### Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> Bind Specifically to *oriT*<sub>LS20</sub>

Electrophoretic Mobility Shift Assays (EMSA) were performed to study the DNA binding properties of Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>. The results presented in **Figure 3** show that both auxiliary proteins bound DNA, and that both bound preferentially to *oriT*<sub>LS20</sub>. Nevertheless, there were distinct differences in binding characteristics between the two proteins. The addition of Aux1<sub>LS20</sub> resulted in the appearance of only one retarded species of *oriT*<sub>LS20</sub>, and even at the highest concentration tested Aux1<sub>LS20</sub> did not bind to the negative control DNA (**Figure 3**, left panel). One retarded *oriT*<sub>LS20</sub> species was also observed for Aux2<sub>LS20</sub> at low concentrations. However, higher Aux2<sub>LS20</sub> concentrations resulted in the appearance of additional shifted species of *oriT*<sub>LS20</sub>. In addition, at higher concentrations Aux2<sub>LS20</sub> bound also to the negative control DNA, and at the highest concentration tested a smear of retarded species was observed (**Figure 3**, right panel). These results show that both proteins bind preferentially to *oriT*<sub>LS20</sub>, but Aux1<sub>LS20</sub> appears to bind *oriT*<sub>LS20</sub> with a higher specificity than Aux2<sub>LS20</sub>.

To delineate further the binding sites of Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> we generated thirteen overlapping 200 bp DNA fragments (F21–F33) covering the *oriT*<sub>LS20</sub> region with a sliding window of 25 bp, and used them in EMSA. The results presented in Supplementary Figure S3 show that Aux1<sub>LS20</sub> bound to fragments F22–F29, which share the 25 bp sequence 5'-CAAATAAATCTGGTACCACGAAAAA-3' located in the 5'



**FIGURE 2 |** Determination of the oligomerization state of the relaxosome proteins Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>. Purified proteins in solution at 12  $\mu$ M were studied by sedimentation velocity (SV) and sedimentation equilibrium (SE). Plots (A, C) are representations of the sedimentation coefficient distribution  $c(s)$  profiles obtained by SV and correspond to Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>, respectively. Concentration gradients obtained by SE assays: (B) Data obtained with Aux1<sub>LS20</sub> (empty circles) and best-fit analysis assuming a protein tetramer (black line), dimer (dashed line) and hexamer (dotted line) species model; (D) Data collected with Aux2<sub>LS20</sub> (empty circles) and best-fitting assuming a protein hexamer (black line), tetramer (dashed line) and octamer (dotted line) species model. Lower part in the SE plots represents the difference between experimental data and estimated values for the best fit to a single species model (residuals).

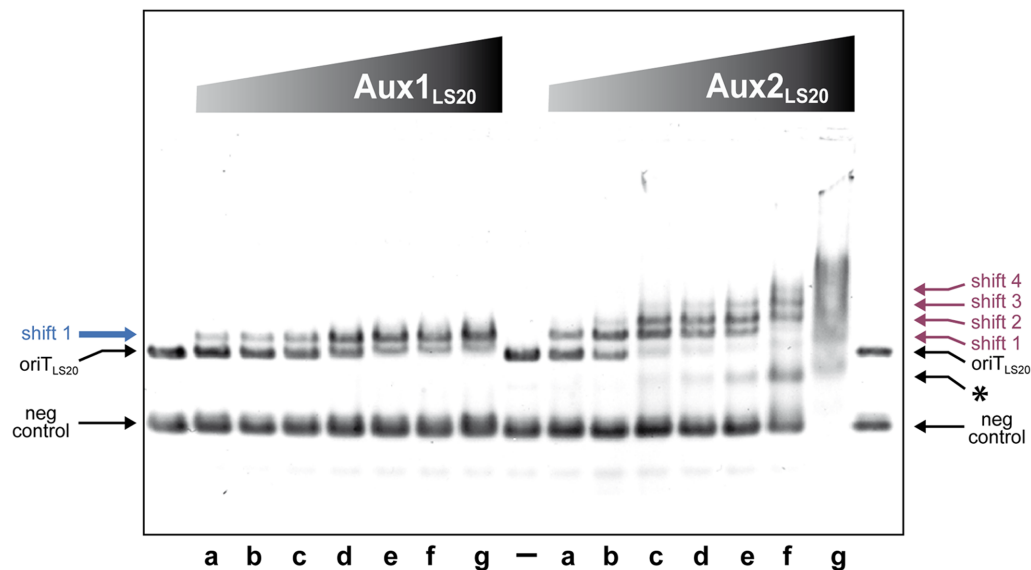
half of *oriT*<sub>LS20</sub>. This sequence contains the inverted repeat 5'-TGGTACCA-3', which could be the binding site of Aux1<sub>LS20</sub>. In the case of Aux2<sub>LS20</sub> retarded species of *oriT*<sub>LS20</sub> with strong and weak intensity were observed for fragments F21–F25 and F26–F28, respectively. No shifts were observed for fragments F29–F33 at the protein concentration used. This shows that Aux2<sub>LS20</sub> binds the 5' half region of *oriT*<sub>LS20</sub> upstream of Aux1<sub>LS20</sub>. The sequence motif 5'-TGTGCAT-3' is present three times in a directed repeated orientation in the 5' half of *oriT*<sub>LS20</sub>. While fragments F21–F25 each contain the three 5'-TGTGCAT-3' motifs, fragment F26 contains only two, and the motif is present only once on fragments F27 and F28. This suggests that the motif 5'-TGTGCAT-3' may be the preferred binding site for Aux2<sub>LS20</sub>. It is worth mentioning that two of the 5'-TGTGCAT-3' motifs are embedded within a larger motif (5'-TTTATGTGCATT-3').

### Over 400 Members of the MOB<sub>L</sub> Family of Relaxase Genes Contain Upstream Genes Encoding Homologs of Aux1<sub>LS20</sub> and/or Aux2<sub>LS20</sub>

Previously, we reported that the pLS20cat-encoded Rel<sub>LS20</sub> constitutes the founding member of a novel, large family of

relaxases that we named MOB<sub>L</sub>, which contained 817 members that were almost exclusively encoded in bacteria belonging to the phylum Firmicutes (Ramachandran et al., 2017). We wanted to know whether other MOB<sub>L</sub> relaxase genes were also preceded by genes encoding putative homologs of Aux1<sub>LS20</sub> and/or Aux2<sub>LS20</sub>. To study this we first determined the current number of MOB<sub>L</sub> relaxase genes, applying the same method as that used in our previous study; i.e., we performed a psi-blastp search of the NCBI nr database using Rel<sub>LS20</sub> as a query. After removing redundant sequences this search now resulted in 1,453 hits that showed high similarity with Rel<sub>LS20</sub> (threshold value  $P = 1e-15$ ). Next, the corresponding DNA accession number of each identified MOB<sub>L</sub> relaxase was retrieved, which was subsequently used to generate a database that contains the accession number of each MOB<sub>L</sub> member together with that of the protein encoded by the gene located upstream and downstream of the relaxase gene. We then performed the same procedures for Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>; i.e., we identified proteins sharing a high level of similarity with Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> and generated databases that contained these accession numbers together with those of the proteins encoded by the flanking genes. Finally, the three databases were crossed to identify those MOB<sub>L</sub> members that are preceded by a gene encoding a putative homolog of Aux1<sub>LS20</sub> and/or Aux2<sub>LS20</sub>. This approach revealed 387 MOB<sub>L</sub>





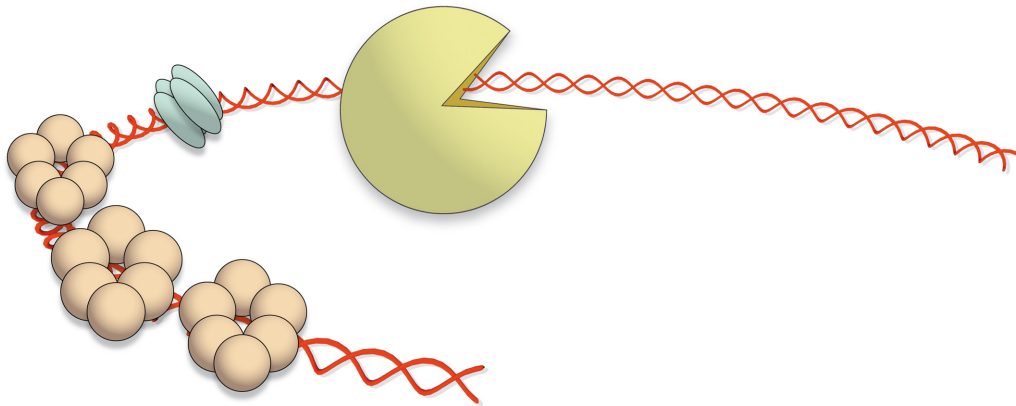
**FIGURE 3 |** Auxiliary protein Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> are DNA binding proteins that preferentially bind to *oriT*<sub>LS20</sub>. DNA binding activity of Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> was analyzed by EMSA. Each lane contained two DNA fragments; one corresponding to *oriT*<sub>LS20</sub> (*oriT*<sub>LS20</sub>, 362 bp, 170 ng), and a control fragment (neg control, 176 bp, 220 ng, corresponding to pLS20 positions 63,774–63,950 of accession number NC\_015148.1). –, loaded without protein. Increasing concentrations of Aux1<sub>LS20</sub> or Aux2<sub>LS20</sub> were prepared using twofold dilution steps, and ranged from 0.09 (lanes “a”) to 5.76  $\mu$ M (lanes “g”). Positions of unbound control DNA (neg control) and *oriT*<sub>LS20</sub> (*oriT*<sub>LS20</sub>) are indicated. The single shifted *oriT*<sub>LS20</sub> species observed for Aux1<sub>LS20</sub> is indicated on the left with a blue arrow. The different shifted *oriT*<sub>LS20</sub> species observed for Aux2<sub>LS20</sub> are indicated on the right with purple arrows, and the shifted species of the control DNA fragment is indicated with an asterisk.

relaxase genes that were preceded by a gene encoding a putative Aux2<sub>LS20</sub> homolog; and of these 87 contained an additional Aux1<sub>LS20</sub> homolog encoding gene upstream. Without exception, the identified MOB<sub>L</sub> relaxase genes having upstream gene(s) encoding putative homologs of Aux1<sub>LS20</sub> and/or Aux2<sub>LS20</sub> are all present in bacteria belonging to the phylum Firmicutes. Although stringent settings were used to identify proteins sharing high similarity with Aux1<sub>LS20</sub> or Aux2<sub>LS20</sub>, this does not automatically imply that the identified proteins will contain a Ribbon-Helix-Helix motif in their N-terminal region, which is a characteristic feature of both Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> (see above). We therefore carried out secondary structure prediction for all the putative Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> homologs identified (see Materials and Methods). The results of these analyses, which are presented in Supplementary Table S4, show that 86 of the 87 (98.9%), and 384 of the 387 (99.2%) putative homologs of Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>, respectively, contain a typical Ribbon-Helix-Helix signature in their N-terminal region, and thereby support the view that they are auxiliary proteins of the corresponding relaxase. In summary, these analyses provide compelling evidence that almost 400 MOB<sub>L</sub> relaxase genes are preceded by a gene encoding an Aux2<sub>LS20</sub> homolog, and that in 87 of these cases this putative auxiliary gene is preceded by another auxiliary gene encoding an Aux1<sub>LS20</sub> homolog. Consequently, pLS20 encoded Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> are the founding members of two families of Ribbon-Helix-Helix type auxiliary proteins that are encoded by Firmicutes bacteria.

## DISCUSSION

In this study we have demonstrated that the pLS20cat genes 56 (*aux1*<sub>LS20</sub>) and 57 (*aux2*<sub>LS20</sub>) encode the auxiliary relaxosome proteins of pLS20cat. Combined with our previously published results (Ramachandran et al., 2017), we have identified the relaxosome module of pLS20cat that includes *oriT*<sub>LS20</sub> and the downstream genes *aux1*<sub>LS20</sub>, *aux2*<sub>LS20</sub>, and *rel*<sub>LS20</sub>. This module is embedded within the large conjugation operon of pLS20cat (Singh et al., 2013). In addition, we have provided strong evidence that Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> constitute the founding member of corresponding families of Ribbon-Helix-Helix type auxiliary proteins whose genes precede a large fraction of the MOB<sub>L</sub> type relaxase genes. Thereby, our results provide a better understanding of the relaxosome components present on Gram+ mobile elements in general and particularly those belonging to the phylum Firmicutes.

The results presented here, together with those obtained previously (Ramachandran et al., 2017), show that *aux1*<sub>LS20</sub> and *aux2*<sub>LS20</sub> encode trans-acting proteins that are essential for conjugation. We also showed that Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> form tetramers and hexamers in solution, respectively, and we detected no interaction between the three pLS20 relaxosome proteins under the conditions tested. We cannot exclude the possibility that they interact when they form a nucleoprotein complex at *oriT*<sub>LS20</sub>. Aux1<sub>LS20</sub> bound with high specificity to a region of 25 bp located about 100 bp upstream of the *nic* site that contains the inverted repeated sequence 5'-TGGTACCA-3'.



**FIGURE 4 |** Schematic model of the nucleoprotein complex formed at *oriT*<sub>LS20</sub>. DNA is shown as a red double helix. The tetrameric Aux1<sub>LS20</sub> unit, the hexameric Aux2<sub>LS20</sub> units and the monomeric Rel<sub>LS20</sub> are depicted as green, orange, and yellow cartoons, respectively. It is well possible that binding of one or more of the auxiliary proteins to *oriT*<sub>LS20</sub> alters the topology of the DNA. This is not taken into account in this simplified model.

The preferred binding site of Aux2<sub>LS20</sub> resulted to be a 140 bp fragment located at the 5' half of *oriT*<sub>LS20</sub> and that contains three times the sequence 5'-TGTGCAT-3'. In our previous study (Ramachandran et al., 2017), we showed that a derivative of *oriT*<sub>LS20</sub> that includes the *nic* site and the binding site for Aux1<sub>LS20</sub>, but lacks the 5'-located 100 bp containing two of the three 5'-TGTGCAT-3' motifs was not functional *in vivo*. The topology of DNA can have a large effect on the binding characteristics of DNA binding proteins and which in turn may affect their function (Gimenes et al., 2008; Fogg et al., 2012). The *oriT* regions of several conjugative plasmids contain an intrinsic bend that is thought to be important for optimal binding and functionality of the relaxosome proteins (for review see, De la Cruz et al., 2010). We have demonstrated that the *oriT*<sub>LS20</sub> region is also intrinsically bent, and that the bend is located in the 5' half of *oriT*<sub>LS20</sub> (Ramachandran et al., 2017), which we show here corresponds to the region where Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> (preferentially) bind. When we combine the results obtained here and in our previous study a picture emerges that is schematically presented in **Figure 4**. Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> bind to the left half of *oriT*<sub>LS20</sub> that is intrinsically bent and we envisage that the formation of this nucleoprotein complex contributes to optimal functioning of Rel<sub>LS20</sub>. In other systems, auxiliary proteins have been described to stimulate relaxase-mediated nicking at *oriT* by recruiting the relaxase to *oriT*, probably by facilitating the relaxase to access the *nic* site, and/or by acting as molecular wedges to melt double-stranded DNA (reviewed in, Alvarez-Martinez and Christie, 2009). Thus, it is conceivable that the auxiliary proteins of pLS20 fulfill similar function(s).

Most of our knowledge on auxiliary proteins is related to those encoded by conjugative plasmids replicating in G<sup>-</sup> bacteria; in particular, the auxiliary proteins of F and related plasmids have been studied in detail at the functional, biochemical and structural levels (for review see, Alvarez-Martinez and Christie, 2009; De la Cruz et al., 2010; Wong et al., 2012). Upon binding, TraY and TraM of plasmid F bent the DNA and

therewith play important roles in organizing the relaxosome complex at *oriT* and influencing the nicking reaction of the relaxase. In addition, they both play a role in gene expression by regulating the activity of their own promoters. TraM also has a key role in delivering the relaxosome to the conjugative pore by interacting with its cognate T4CP (Wong et al., 2011; Peng et al., 2014). Future studies are needed to determine whether the auxiliary proteins of pLS20 fulfill similar functions to those of F, although it is doubtful that Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> play a role in gene regulation due to the different genetic organization. In the case of F, the monocistronic *traM* gene is located directly downstream of its *oriT*. *TraM* is followed by another monocistronic gene, *traJ*, which in turn is followed by a large multicistronic operon in which *traY* is the first gene (Zatyka and Thomas, 1998). In the case of pLS20, though, the relaxosome genes are embedded within the large conjugation operon and are under the control of the main conjugation promoter P<sub>c</sub> that is located almost 26 kbp upstream of *aux1*<sub>LS20</sub> (Singh et al., 2013; Ramachandran et al., 2014). At present, we cannot fully exclude the possibility that the relaxosome genes of pLS20cat are controlled by an additional promoter that is regulated by Aux1<sub>LS20</sub> or Aux2<sub>LS20</sub>. RNAseq data showed, however, that repression of the main conjugation promoter results in silencing of the relaxosome genes, as well as other genes in the conjugation operon of pLS20cat (Singh et al., 2013).

Far less is known about auxiliary proteins encoded by conjugative plasmids of Gram<sup>+</sup> origin. The monomeric Helix-Turn-Helix protein TraN of the *Enterococcus faecalis* conjugative plasmid pIP501 binds to its *oriT* region, which suggested that it might be an auxiliary protein of pIP501. However, recent results revealed that *traN* is not essential for conjugation, and it is now believed that it may be a repressor of conjugation by regulating either the expression of the conjugation operon or activity of the relaxase TraA (Goessweiner-Mohr et al., 2014; Grohmann et al., 2016).

The auxiliary proteins PcfF encoded by the *E. nterococcus faecalis* plasmid pCF10, and LtrF of *Lactococcus lactis* plasmid pRS01 share 47% sequence identity. As far as we know, these are the only auxiliary proteins encoded by conjugative plasmids of Gram<sup>+</sup> origin that have been studied in some detail (Chen et al., 2007, 2008). The *pcfF* and *ltrF* genes are essential for conjugation and purified PcfF and LtrF bind their cognate *oriT*s. Moreover, evidence supports a model in which PcfF recruits the relaxase PcfG to *oriT*, and that PcfF, probably in conjunction with the relaxase PcfG, interacts with its cognate T4CP and hence plays an important role in delivering the relaxosome to the conjugative pore.

Several auxiliary proteins of conjugative plasmids of Gram<sup>−</sup> origin are described to contain a RHH motif. These include, TraY and TraM of F plasmid, TrwA of R388, VirC2 of *Agrobacterium tumefaciens*, NikA of R64, TraJ of RP4, MobC of RSF1010, MbeC of ColE1, MobC of RA3 (Bowie and Sauer, 1990; Zhang and Meyer, 1997; Moncalian and De la Cruz, 2004; Ragonese et al., 2007; Yoshida et al., 2008; Lu et al., 2009; Varsaki et al., 2009; Godziszewska et al., 2016). For some of them structure-based mutational analyses have demonstrated the importance of the RHH motif in *oriT* binding as well as relaxase recruitment (Yoshida et al., 2008; Lu et al., 2009). Interestingly, Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> are also predicted to contain an RHH DNA-binding domain in their N-terminal region (Supplementary Figure S1). In addition, our *in silico* analyses predict that the auxiliary PcfF and LtrF proteins of Gram<sup>+</sup> *E. faecalis* pCF10 and *L. lactis* pRS01 plasmids, respectively, also contain an RHH motif in their N-terminal region (our unpublished results). The presence of a likely RHH motif in Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> is therefore in line with the conclusion that they are auxiliary proteins. More importantly, the observation that the auxiliary proteins encoded by plasmids pLS20, pRS01, and pCF10, replicating in Gram<sup>+</sup> bacteria, all contain a predicted RHH motif indicates that this is a conserved motif in auxiliary proteins encoded by CEs of both Gram<sup>−</sup> and Gram<sup>+</sup> origin, and suggests that auxiliary proteins share a common ancestor. We have made use of this feature, combined with the genetic organization, to identify putative auxiliary genes located upstream of the MOB<sub>L</sub> type relaxase genes that encode homologs of Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub>. This strategy resulted in the identification of about 400 and 90 genes encoding homologs of Aux2<sub>LS20</sub> and Aux1<sub>LS20</sub>, respectively; 99.2% (Aux2<sub>LS20</sub>) and 98.9% (Aux1<sub>LS20</sub>) of these homologs were predicted to contain a Ribbon-Helix-Helix motif in their N-terminal region. These results reinforce therefore the view that an N-terminal Ribbon-Helix-Helix DNA binding motif is a characteristic feature of auxiliary relaxosome proteins. In addition, these data showed that Aux1<sub>LS20</sub> and Aux2<sub>LS20</sub> are the founding members of two families of auxiliary proteins whose genes are genetically linked to a MOB<sub>L</sub> type relaxase gene. In summary, we have demonstrated that pLS20cat genes 56 (*aux1<sub>LS20</sub>*) and 57 (*aux2<sub>LS20</sub>*) encode the

auxiliary proteins of pLS20 that are essential for conjugation, and that they form the founding members of families of auxiliary relaxosome proteins that are encoded in Firmicutes bacteria.

## AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct experimental and/or intellectual contribution to the work. AM-A, J-AH, GR, CG-C, DG-A, and JV-C generated all plasmids and strains, purified proteins and executed all the experiments except the ultracentrifugation studies, which were performed by JL-O and CA. DA performed *in silico* analyses contributed to the general design and analyses of the results. LW and WM designed the experimental plan and were principally responsible for analyzing the results and writing the paper. WM supervised AM-A, J-AH, GR, CG-C, DG-A and JV-C.

## FUNDING

Work in the Meijer lab was funded by grants Bio2013-41489-P and BIO2016-77883-C2-1-P of the Ministry of Economy and Competitiveness of the Spanish Government to WM, which also funded AM-A, CG-C, and JV-C. Part of the economic support of the two aforementioned grants was provided by the “Agencia Estatal de Investigación (AEI)” and “Fondo Europeo de Desarrollo Regional (FEDER).” This research was also supported by institutional grants from the “Fundación Ramón Areces” and “Banco de Santander” to the Centro de Biología Molecular “Severo Ochoa”. LW’s work was supported by Wellcome Trust grant WT098374AIA to Jeff Errington. JL-O and CA were supported by grant BFU2014-52070-C2-2-P of the Ministry of Science and Innovation to CA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. J-AH received a State Scholarship Fund from the China Scholarship Council.

## ACKNOWLEDGMENTS

We thank Jose Belio for help with preparing the Figures, and Margarita Salas and Jeff Errington for their support on our work. We also want to acknowledge helpful discussion with other lab members, and particularly want to mention Praveen K. Singh.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02138/full#supplementary-material>

## REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., et al. (2005). Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* 272, 5101–5109. doi: 10.1111/j.1742-4658.2005.004945.x
- Alvarez-Martinez, C. E., and Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* 73, 775–808. doi: 10.1128/MMBR.00023-09
- Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* 277, 819–827. doi: 10.1098/rspb.2009.1679
- Bowie, J. U., and Sauer, R. T. (1990). TraY proteins of F and related episomes are members of the Arc and Mnt repressor family. *J. Mol. Biol.* 211, 5–6.
- Bron, S., Peijnenburg, A., Peeters, B. P. H., Haima, P., and Venema, G. (1989). “Cloning and plasmid (in)stability in *Bacillus subtilis*,” in *Genetic Transformation and Expression*, eds L. O. Butler, C. R. Harwood, and B. E. B. Moseley (Andover: Intercept Ltd.), 205–219.
- Chen, Y., Staddon, J. H., and Dunny, G. M. (2007). Specificity determinants of conjugative DNA processing in the *Enterococcus faecalis* plasmid pCF10 and the *Lactococcus lactis* plasmid pRS01. *Mol. Microbiol.* 63, 1549–1564.
- Chen, Y., Zhang, X., Manias, D., Yeo, H. J., Dunny, G. M., and Christie, P. J. (2008). *Enterococcus faecalis* PcfC, a spatially localized substrate receptor for type IV secretion of the pCF10 transfer intermediate. *J. Bacteriol.* 190, 3632–3645. doi: 10.1128/JB.01999-07
- Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433. doi: 10.1128/MMBR.00016-10
- De la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.* 34, 18–40. doi: 10.1111/j.1574-6976.2009.00195.x
- Del Solar, G., Hernandez-Arriaga, A. M., Gomis-Ruth, F. X., Coll, M., and Espinosa, M. (2002). A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number. *J. Bacteriol.* 184, 4943–4951.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Fogg, J. M., Randall, G. L., Pettitt, B. M., Sumners de, W. L., Harris, S. A., and Zechiedrich, L. (2012). Bullied no more: when and how DNA shoves proteins around. *Q. Rev. Biophys.* 45, 257–299. doi: 10.1017/S0033583512000054
- Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O., and Dantas, G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337, 1107–1111. doi: 10.1126/science.1220761
- Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732.
- Gimenes, F., Takeda, K. I., Fiorini, A., Gouveia, F. S., and Fernandez, M. A. (2008). Intrinsically bent DNA in replication origins and gene promoters. *Genet. Mol. Res.* 7, 549–558.
- Godziszewska, J., Moncalian, G., Cabezas, M., Bartosik, A. A., De la Cruz, F., and Jagura-Burdzy, G. (2016). Concerted action of NIC relaxase and auxiliary protein MobC in RA3 plasmid conjugation. *Mol. Microbiol.* 101, 439–456. doi: 10.1111/mmi.13401
- Goessweiner-Mohr, N., Arends, K., Keller, W., and Grohmann, E. (2013). Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid* 70, 289–302. doi: 10.1016/j.plasmid.2013.09.005
- Goessweiner-Mohr, N., Eder, M., Hofer, G., Fercher, C., Arends, K., Birner-Gruenberger, R., et al. (2014). Structure of the double-stranded DNA-binding type IV secretion protein TraN from *Enterococcus*. *Acta Crystallogr. D. Biol. Crystallogr.* 70, 2376–2389. doi: 10.1107/S1399004714014187
- Gomis-Ruth, F. X., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., et al. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.* 17, 7404–7415.
- Grohmann, E., Goessweiner-Mohr, N., and Brantl, S. (2016). DNA-binding proteins regulating pIP501 transfer and replication. *Front. Mol. Biosci.* 3:42. doi: 10.3389/fmolb.2016.00042
- Itaya, M., Sakaya, N., Matsunaga, S., Fujita, K., and Kaneko, S. (2006). Conjugational transfer kinetics of pLS20 between *Bacillus subtilis* in liquid medium. *Biosci. Biotechnol. Biochem.* 70, 740–742.
- Kopeck, J., Bergmann, A., Fritz, G., Grohmann, E., and Keller, W. (2005). TraA and its N-terminal relaxase domain of the Gram-positive plasmid pIP501 show specific *oriT* binding and behave as dimers in solution. *Biochem. J.* 387, 401–409.
- Lee, C. A., and Grossman, A. D. (2007). Identification of the origin of transfer (*oriT*) and DNA relaxase required for conjugation of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *J. Bacteriol.* 189, 7254–7261.
- Lu, J., den Dulk-Ras, A., Hooykaas, P. J., and Glover, J. N. (2009). *Agrobacterium tumefaciens* VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9643–9648. doi: 10.1073/pnas.0812199106
- Mazel, D., and Davies, J. (1999). Antibiotic resistance in microbes. *Cell Mol. Life Sci.* 56, 742–754.
- Moncalian, G., and De la Cruz, F. (2004). DNA binding properties of protein TraA, a possible structural variant of the Arc repressor superfamily. *Biochim. Biophys. Acta* 1701, 15–23.
- Norman, A., Hansen, L. H., and Sorensen, S. J. (2009). Conjugative plasmids: vessels of the communal gene pool. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 364, 2275–2289. doi: 10.1098/rstb.2009.0037
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.
- Penders, J., Stobberingh, E. E., Savellkoul, P. H., and Wolfs, P. F. (2013). The human microbiome as a reservoir of antimicrobial resistance. *Front. Microbiol.* 4:87. doi: 10.3389/fmicb.2013.00087
- Peng, Y., Lu, J., Wong, J. J., Edwards, R. A., Frost, L. S., and Mark Glover, J. N. (2014). Mechanistic basis of plasmid-specific DNA binding of the F plasmid regulatory protein, TraM. *J. Mol. Biol.* 426, 3783–3795. doi: 10.1016/j.jmb.2014.09.018
- Pluta, R., Boer, D. R., Lorenzo-Diaz, F., Russi, S., Gomez, H., Fernandez-Lopez, C., et al. (2017). Structural basis of a histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6526–E6535. doi: 10.1073/pnas.1702971114
- R Core Team (2017). *R: A Language, and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ragonese, H., Haisch, D., Villareal, E., Choi, J. H., and Matson, S. W. (2007). The F plasmid-encoded TraM protein stimulates relaxosome-mediated cleavage at *oriT* through an interaction with TraI. *Mol. Microbiol.* 63, 1173–1184.
- Ramachandran, G., Miguel-Arribas, A., Abia, D., Singh, P. K., Crespo, I., Gago-Cordoba, C., et al. (2017). Discovery of a new family of relaxases in Firmicutes bacteria. *PLOS Genet.* 13:e1006586. doi: 10.1371/journal.pgen.1006586
- Ramachandran, G., Singh, P. K., Luque-Ortega, J. R., Yuste, L., Alfonso, C., Rojo, F., et al. (2014). A complex genetic switch involving overlapping divergent promoters and DNA looping regulates expression of conjugation genes of a gram-positive plasmid. *PLOS Genet.* 10:e1004733. doi: 10.1371/journal.pgen.1004733
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., et al. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29, 2994–3005.
- Schildbach, J. F., Karzai, A. W., Raumann, B. E., and Sauer, R. T. (1999). Origins of DNA-binding specificity: role of protein contacts with the DNA backbone. *Proc. Natl. Acad. Sci. U.S.A.* 96, 811–817.
- Singh, P. K., Ramachandran, G., Duran-Alcalde, L., Alonso, C., Wu, L. J., and Meijer, W. J. (2012). Inhibition of *Bacillus subtilis* natural competence by a native, conjugative plasmid-encoded comK repressor protein. *Environ. Microbiol.* 14, 2812–2825. doi: 10.1111/j.1462-2920.2012.02819.x
- Singh, P. K., Ramachandran, G., Ramos-Ruiz, R., Peiro-Pastor, R., Abia, D., Wu, L. J., et al. (2013). Mobility of the native *Bacillus subtilis* conjugative plasmid pLS20 is regulated by intercellular signaling. *PLOS Genet.* 9:e1003892. doi: 10.1371/journal.pgen.1003892
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and De la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452.



- Sommer, M. O., Church, G. M., and Dantas, G. (2010). The human microbiome harbors a diverse reservoir of antibiotic resistance genes. *Virulence* 1, 299–303. doi: 10.4161/viru.1.4.12010
- Sommer, M. O., Dantas, G., and Church, G. M. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325, 1128–1131. doi: 10.1126/science.1176950
- Tanaka, T., Kuroda, M., and Sakaguchi, K. (1977). Isolation and characterization of four plasmids from *Bacillus subtilis*. *J. Bacteriol.* 129, 1487–1494.
- Thoma, L., and Muth, G. (2016). Conjugal DNA-transfer in *Streptomyces*, a mycelial organism. *Plasmid* 87–88, 1–9. doi: 10.1016/j.plasmid.2016.09.004
- Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721.
- Varsaki, A., Moncalian, G., Garcillan-Barcia, M. P., Drainas, C., and De la Cruz, F. (2009). Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. *J. Bacteriol.* 191, 1446–1455. doi: 10.1128/JB.01342-08
- Wang, S., Li, W., Liu, S., and Xu, J. (2016). RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* 44, W430–W435. doi: 10.1093/nar/gkw306
- Waters, V. L. (1999). Conjugal transfer in the dissemination of beta-lactam and aminoglycoside resistance. *Front. Biosci.* 4:D433–D456.
- Wong, J. J., Lu, J., Edwards, R. A., Frost, L. S., and Glover, J. N. (2011). Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. *Nucleic Acids Res.* 39, 6775–6788. doi: 10.1093/nar/gkr296
- Wong, J. J., Lu, J., and Glover, J. N. (2012). Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Mol. Microbiol.* 85, 602–617. doi: 10.1111/j.1365-2958.2012.08131.x
- Yoshida, H., Furuya, N., Lin, Y. J., Guntert, P., Komano, T., and Kainosho, M. (2008). Structural basis of the role of the NikA ribbon-helix-helix domain in initiating bacterial conjugation. *J. Mol. Biol.* 384, 690–701. doi: 10.1016/j.jmb.2008.09.067
- Zatyka, M., and Thomas, C. M. (1998). Control of genes for conjugative transfer of plasmids and other mobile elements. *FEMS Microbiol. Rev.* 21, 291–319.
- Zhang, S., and Meyer, R. (1997). The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. *Mol. Microbiol.* 25, 509–516.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with several of the authors JL-O and CA.

Copyright © 2017 Miguel-Arribas, Hao, Luque-Ortega, Ramachandran, Val-Calvo, Gago-Córdoba, González-Álvarez, Abia, Alfonso, Wu and Meijer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Bad Phages in Good Bacteria: Role of the Mysterious *orf63* of $\lambda$ and Shiga Toxin-Converting $\Phi 24_B$ Bacteriophages

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Radosław Pluta,  
International Institute of Molecular and  
Cell Biology in Warsaw (IIMCB),  
Poland

Ramon Diaz Orejas,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain  
Dhruva Chatteraj,  
National Institutes of Health,  
United States

### \*Correspondence:

Alicja Węgrzyn  
alicja.wegrzyn@biol.ug.edu.pl

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 June 2017

**Accepted:** 08 August 2017

**Published:** 25 August 2017

### Citation:

Dydecka A, Bloch S, Rizvi A, Perez S,  
Nejman-Falenczyk B, Topka G,  
Gasior T, Necel A, Węgrzyn G,  
Donaldson LW and Węgrzyn A (2017)  
Bad Phages in Good Bacteria: Role of  
the Mysterious *orf63* of  $\lambda$  and Shiga  
Toxin-Converting  $\Phi 24_B$   
Bacteriophages.  
Front. Microbiol. 8:1618.  
doi: 10.3389/fmicb.2017.01618

Aleksandra Dydecka<sup>1†</sup>, Sylwia Bloch<sup>1†</sup>, Ali Rizvi<sup>2</sup>, Shaili Perez<sup>2</sup>,  
Bożena Nejman-Falenczyk<sup>1</sup>, Gracja Topka<sup>1</sup>, Tomasz Gasior<sup>3</sup>, Agnieszka Necel<sup>1</sup>,  
Grzegorz Węgrzyn<sup>1</sup>, Logan W. Donaldson<sup>2</sup> and Alicja Węgrzyn<sup>3\*</sup>

<sup>1</sup> Department of Molecular Biology, Faculty of Biology, University of Gdańsk, Gdańsk, Poland, <sup>2</sup> Department of Biology, York University, Toronto, ON, Canada, <sup>3</sup> Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

Lambdoid bacteriophages form a group of viruses that shares a common schema of genome organization and lifecycle. Some of them can play crucial roles in creating the pathogenic profiles of *Escherichia coli* strains. For example, Shiga toxin-producing *E. coli* (STEC) acquired *stx* genes, encoding Shiga toxins, via lambdoid prophages (Stx phages). The results obtained so far present the evidence for the relation between the *exo-xis* region of the phage genome and lambdoid phage development, however molecular mechanisms of activities of the *exo-xis* genes' products are still unknown. In view of this, we decided to determine the influence of the uncharacterized open reading frame *orf63* of the *exo-xis* region on lambdoid phages development using recombinant prophages,  $\lambda$  and Stx phage  $\Phi 24_B$ . We have demonstrated that *orf63* codes for a folded protein, thus, it is a functional gene. NMR spectroscopy and analytical gel filtration were used to extend this observation further. From backbone chemical shifts, Orf63 is oligomeric in solution, likely a trimer and consistent with its small size (63 aa.), is comprised of two helices, likely intertwined to form the oligomer. We observed that the deletion of phage *orf63* does not impair the intracellular lambdoid phage lytic development, however delays the time and decreases the efficiency of prophage induction and in consequence results in increased survival of *E. coli* during phage lytic development. Additionally, the deletion of phage *orf63* negatively influences expression of the major phage genes and open reading frames from the *exo-xis* region during prophage induction with hydrogen peroxide. We conclude, that lambdoid phage *orf63* may have specific functions in the regulation of lambdoid phages development, especially at the stage of the lysis vs. lysogenization decision. Besides, *orf63* probably participates in the regulation of the level of expression of essential phage genes and open reading frames from the *exo-xis* region during prophage induction.

**Keywords:** Shiga toxin-producing *Escherichia coli* (STEC), lambdoid bacteriophages, lytic development, *exo-xis* region, open reading frames

## INTRODUCTION

The significance of Shiga toxin-producing *E. coli* (STEC) as a public health problem was first recognized in 1982 during an investigation of an outbreak of hemorrhagic colitis associated with consumption of contaminated hamburgers (Riley et al., 1983). Since then, STEC strains have been implicated in many outbreaks of diarrhea world-wide. Quite recently (2011), the Shiga toxin-producing *E. coli* serotype O104:H4 was responsible for a serious epidemic outbreak in Germany (Bloch et al., 2012; Muniesa et al., 2012). STEC pathogens can cause serious food poisoning with bloody diarrhea in humans (Nataro and Kaper, 1998). Their main virulence factors are Shiga toxins, encoded by *stx* genes located in genomes of bacteriophages which occur in bacteria as prophages (Mizutani et al., 1999). These bacteriophages are called Shiga toxin-converting or Stx, for short, and belong to the lambdoid family of phages (Schmidt, 2001). All phages within this group indicate high similarities in the lifecycle and genomic organization to bacteriophage  $\lambda$ , the most reviewed member of this family (Wegrzyn et al., 2012). In the prophage state, most of phage genes, including *stx* genes, are not transcribed due to inhibition caused by the phage *cI* repressor. As a consequence, Shiga toxins are not produced under such conditions. Expression of *stx* as well as other phage genes occurs effectively only after prophage induction. In most cases, this process requires activation of the RecA-dependent bacterial S.O.S. response which is provoked by factors causing appearance of single-stranded DNA fragments. Activated RecA protein stimulates cleavage of the S.O.S. regulon repressor, the LexA protein, and the *cI* phage repressor. Prophage induction and subsequent phage lytic development lead to production of progeny phage particles and Shiga toxins, followed by their release from the lysed cell (Licznarska et al., 2016b). In the regulation of the lysis-vs. -lysogenization decision after infection of the host cell by a bacteriophage, both phage- and host-encoded proteins play important roles (for a review, see Wegrzyn et al., 2012). Among environmental factors influencing the decision, the crucial are temperature, nutrients availability and multiplicity of infection (m.o.i.). Lytic growth is supported by high temperature, rich medium and high m.o.i., while low temperature, starvation and low m.o.i. favor lysogenization. At the molecular level, the major players supporting lytic and lysogenic pathways are Cro and *cI* proteins, respectively. They are transcriptional regulators, and Cro represses expression of the *cI* gene, whereas *cI* downregulates transcription from two major “lytic” promoters (pL and pR, which provide mRNAs for *cro* and other “lytic” genes, encoding proteins involved in all processes during production of phage progeny) while stimulating its own expression by activation of the pM promoter. Thus, the result of the competition between Cro and *cI* is crucial for choosing one of the alternative developmental pathways. Since shortly after infection no *cI* protein is present, another transcription regulator, the *cII* protein (whose gene is transcribed from pR), is a key player in this game. This protein activates the second promoter for *cI* expression, pE. Therefore, *cII* activity decides on the Cro or *cI* predominance. In fact, *cII* is a subject of various regulatory mechanisms acting in response to different

environmental conditions, including those playing major roles in the lysis-vs. -lysogenization decision (see Wegrzyn et al., 2012, for details).

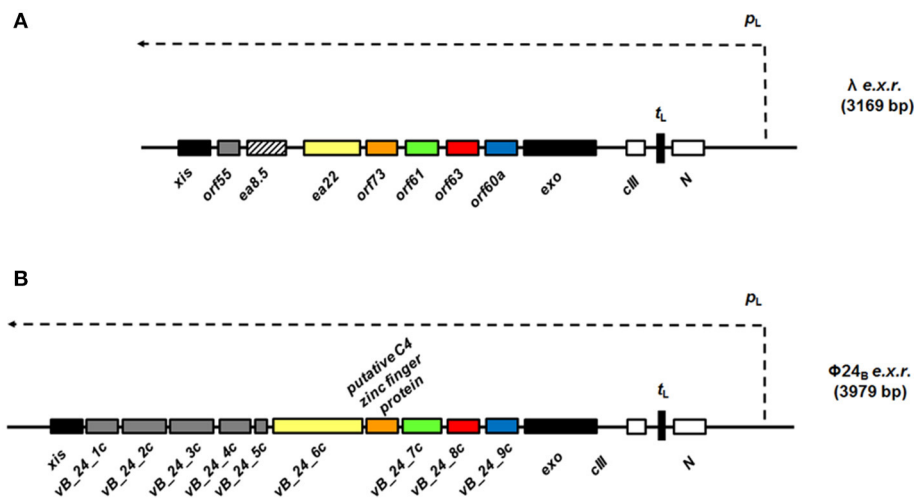
An evolutionarily conserved region of lambdoid bacteriophage genome, located between *exo* and *xis* genes (so called “the *exo-xis* region”), contains several genes and open reading frames (Figure 1). Quite surprisingly, until recently, the role of this region in bacteriophage development was almost completely unknown. Recent studies indicated that overexpression of genes from the *exo-xis* region’ impaired lysogenization of *E. coli* by bacteriophage  $\lambda$  (Loś et al., 2008b) and enhanced induction of prophages  $\lambda$  and  $\Phi 24_B$  (one of Shiga toxin-converting phages) (Bloch et al., 2013). The Ea8.5 protein, encoded by a gene located in the *exo-xis* region, contains a fused homeodomain/zinc-finger fold (Kwan et al., 2013), which suggest a regulatory role for this protein. Interestingly, prophage induction with mitomycin C or hydrogen peroxide caused different expression patterns of genes from the *exo-xis* region; such differences were observed in both phages,  $\lambda$  and  $\Phi 24_B$  (Bloch et al., 2014). Moreover, phages with deletions in the *exo-xis* region responded to the oxidative stress in a different manner relative to wild-type phages (Licznarska et al., 2016a). Therefore, it is important to determine structures and functions of particular proteins encoded in the *exo-xis* region.

In this work, we have focused on *orf63*. In our preliminary experiments with mutants in particular genes and ORFs from the *exo-xis* region, deletion of *orf63* gave one of the strongest effects (Licznarska et al., 2016a). Moreover, the transcription factor YqhC has been recognized as a potential partner for interaction with Orf63 in yeast two-hybrid study of phage-host interactions (Blasche et al., 2013). Therefore, we decided to investigate structure and functions of *orf63* and its product, the Ofr63 protein, in more detail.

## MATERIALS AND METHODS

### The *orf63* Gene Expression and Protein Purification

A codon-optimized *orf63* gene (NCBI ID: 2703507) for high level expression *E. coli* was synthesized by ATUM (Menlo Park, CA) and supplied in plasmid pD441-NH for direct transformation of a BL21 host strain (Novagen). Amino terminal 6xHis and Flag (DYKDDDDK) tags were included to facilitate affinity purification and detection. Milligram quantities of isotopically labeled 6xHis-Flag-Orf63 for NMR spectroscopy were obtained from a 1.0 L fermentation in a minimal medium containing 1 g  $^{15}\text{NH}_4\text{Cl}$ , 3 g of  $^{13}\text{C}$ -glucose, and 1 g of  $^{15}\text{N}$ - $^{13}\text{C}$  Celtone algal extract (CIL; Cambridge, MA). The cell pellet was dissolved in T300 buffer (20 mM Tris-HCl, 300 mM NaCl, 0.05%  $\text{NaN}_3$ ) and lysed by French press and sonication. The Orf63 protein purified from the bacterial soluble fraction by Nickel-NTA affinity chromatography (Qiagen) that included a 10 mM imidazole wash step and a 20 mM EDTA elution step, all in T300 buffer. A subsequent gel filtration chromatography step (Sephacryl-100, HiLoad 16/60; GE Life Sciences) was employed to further purify the Orf63 protein and exchange it into NMR spectroscopy buffer (5 mM Tris-HCl, 0.15 M NaCl, 0.05%  $\text{NaN}_3$ ).



**FIGURE 1** | Maps of genes and open reading frames (ORFs) from the region located between *exo* and *xis* genes (black rectangles), the *exo-xis* region or *e.x.r.*, of lambdoid bacteriophages:  $\lambda$  (**A**) and  $\Phi 24_B$  (**B**). In the case of phage  $\lambda$  (**A**), the *exo-xis* region consists of two recognized genes: *ea22* and *ea8.5*, and five additional ORFs, named: *orf60a*, *orf63*, *orf61*, *orf73*, and *orf55*, which expression is under control of the  $p_L$  promoter (thin dashed arrow). Comparatively, the *exo-xis* region of phage  $\Phi 24_B$  (**B**) contains additional ORFs (gray rectangles), but there is no homolog of the *ea8.5* gene of phage  $\lambda$  (**A**, rectangle with black stripes). Note, that some ORFs from the *exo-xis* region of phage  $\Phi 24_B$  (**B**): *vb\_24B\_9c* (blue rectangles), *vb\_24B\_8c* (red rectangles), *vb\_24B\_7c* (green rectangles), putative C4 zinc finger protein (orange rectangles), and *vb\_24B\_6c* (yellow rectangles) are homologs of phage  $\lambda$  *orf60a*, *orf63*, *orf61*, *orf73*, *ea22* (**B**), respectively. In spite of the differences in composition of both  $\lambda$  and  $\Phi 24_B$  *exo-xis* regions, attention needs to be paid to highly conserved sequences of the *orf60a-orf73* regions among lambdoid bacteriophages ( $\geq 70\%$  nucleotide and amino acid sequence identity) (Bloch et al., 2013). The regulatory genes: *N* and *cIII* are marked as white rectangles and  $t_L$  terminator is indicated as black vertical rectangle.

## Gel Filtration Assay to Estimate Apparent Molecular Weight

Chromatograms of eleven proteins on the same gel filtration column used to purify Orf63 were compiled to produce a standard curve describing the relationship between retention volume and molecular weight. Specific proteins used for the standard curve included the HACSI SH3 domain (10.6 kDa), the AIDA1 PTB domain (22.0 kDa), the monomeric and dimeric states CASKIN2 SAM domain tandem (20.2 / 40.5 kDa), the CASKIN2 SAM1 domain (10.3 kDa), the SHP2 adaptor SH2 domain (14.1 kDa), the AIDA1 SAM domain tandem (16.4 kDa), two deletion mutants of the La RRM domain (14.6 / 15.9 kDa), the Crk2 adaptor SH2 domain (15.5 kDa), dimeric glutathione S-transferase (52 kDa), and calmodulin (18.8 kDa).

## NMR Spectroscopy

A 0.5 mM sample of  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled Orf63 was prepared for NMR spectroscopy in NMR buffer supplemented with 10%  $\text{D}_2\text{O}$ . All experiments were performed at 310 K using a Bruker Avance 700 MHz NMR spectrometer equipped with a cryogenically cooled 5 mm probe at the York University Life Sciences Building Central Facility. Backbone (HN, N, CA, CB, C') assignments were achieved using a set of conventional triple resonance experiments (HNCA, HNCACB, CBCAcoNH, HNCO, HNcaCO) incorporating sparse sampling for the optimum sensitivity and resolution. Datasets were processed with NMRpipe (Delaglio et al., 1995) and istHMS (Hyberts et al., 2012) and interpreted with CCPN Analysis (Skinner et al., 2015).

## Bacteria, Bacteriophages and Plasmids

The *E. coli* strains, bacteriophages and plasmids used in *in vivo* work are presented in Table 1. Work with these strains was approved by the Ministry of Environment (decision no. 189/2016). The *E. coli* lysogens were obtained using the following lambdoid phages:  $\lambda$ ,  $\lambda\Delta\text{orf63}$ ,  $\Phi 24_B$  or  $\Phi 24_B\Delta\text{orf63}$  (Bloch et al., 2013; Licznarska et al., 2016b). In the first step, phage lysates were prepared. Bacterial cultures were grown at 37°C to  $A_{600} = 0.1$ . Then, mitomycin C (Sigma—Aldrich) was added to all flasks to a final concentration of 1  $\mu\text{g/ml}$ . The incubation with shaking was continued for about 12 h. To obtain lysates, bacterial debris were centrifuged ( $2,000 \times g$  for 10 min at 4°C) and supernatants were filtered through the 0.22- $\mu\text{m}$ -pore-size filters (Sigma—Aldrich). In the next step, the lysogenization procedure was carried out. Briefly, *E. coli* strain C600 was cultivated at 37°C to  $A_{600} = 0.2$ . Then, 4 ml of bacterial culture was centrifuged ( $2,000 \times g$  for 5 min at RT), the pellet was washed with TCM buffer (10 mM Tris-HCl, 10 mM  $\text{MgSO}_4$ , 10 mM  $\text{CaCl}_2$ , pH 7.2; Sigma—Aldrich) and suspended in LB medium (Sigma—Aldrich) supplemented with  $\text{MgSO}_4$  (phages  $\lambda$  and  $\lambda\Delta\text{orf63}$ ) or with  $\text{MgSO}_4$  and  $\text{CaCl}_2$  (phages  $\Phi 24_B$  and  $\Phi 24_B\Delta\text{orf63}$ ) to a final concentration of 10 mM. Bacteriophages were added to the suspensions to m.o.i. of 10. Following incubation at 37°C, the mixtures were spread on LB agar plates. After overnight incubation at 37°C, bacterial colonies were tested for the presence of prophages by using UV irradiation (this procedure is described in detail in the next section). For construction of the plasmid pSB\_orf63\_ $\lambda$ , nucleotide sequence of *orf63* from phage  $\lambda$  was amplified by PCR with primers: F $\lambda$ orf63\_EcoRI (5'GGA GAA TTC GGC TGT ATG CAC AAA



TABLE 1 | Bacterial strains, bacteriophages and plasmids used for *in vivo* experiments.

<i>E. coli</i> strains, bacteriophages or plasmids	Relevant genotype or description	References
<b><i>E. coli</i> LABORATORY STRAINS</b>		
MG1655	F– $\lambda$ – <i>ilvG rfb-50 rph-1</i>	Jensen, 1993
MG1655 ( $\lambda$ )	MG1655 bearing $\lambda$ prophage	Bloch et al., 2013
MG1655 ( $\lambda\Delta orf63$ )	MG1655 bearing $\lambda$ prophage with deletion of <i>orf63</i>	Licznarska et al., 2016a
MG1655 ( $\Phi24_B$ )	MG1655 bearing $\Phi24_B$ prophage	Bloch et al., 2013
MG1655 ( $\Phi24_B\Delta orf63$ )	MG1655 bearing $\Phi24_B$ prophage with deletion of <i>vb_24B_8c</i> , the homolog of $\lambda orf63$	Licznarska et al., 2016a
C600	F– <i>tonA21 thi-1 thr-1 leuB6 lacY1 glnV44 rfbC1 fhuA1</i> $\lambda^-$	Appleyard, 1954
C600 ( $\lambda$ )	C600 bearing $\lambda$ prophage	This study, by lysogenization
C600 ( $\lambda\Delta orf63$ )	C600 bearing $\lambda$ prophage with deletion of <i>orf63</i>	This study, by lysogenization
C600 ( $\Phi24_B$ )	C600 bearing $\Phi24_B$ prophage	This study, by lysogenization
C600 ( $\Phi24_B\Delta orf63$ )	C600 bearing $\Phi24_B$ prophage with deletion of <i>vb_24B_8c</i> , the homolog of $\lambda orf63$	This study, by lysogenization
<b>BACTERIOPHAGES</b>		
$\lambda$	carries a frameshift mutation relative to Ur-lambda	Hendrix and Duda, 1992
$\lambda\Delta orf63$	$\Phi24_B$ phage with deletion of <i>orf63</i>	Licznarska et al., 2016a
$\Phi24_B$	$\lambda stx2::cat$	Allison, 2003
$\Phi24_B\Delta orf63$	$\Phi24_B$ phage with deletion of <i>vb_24B_8c</i> , the homolog of $\lambda orf63$	Licznarska et al., 2016a
<b>PLASMIDS</b>		
pUC18	<i>ori</i> pMB1 (pBR322 derivative), <i>bla</i> , Amp <sup>R</sup>	Thermo fisher scientific
pSB_orf63_ $\lambda$	as pUC18 but bearing the <i>orf63</i> from phage $\lambda$	This study
pSB_orf63_ $\Phi24_B$	as pUC18 but bearing the <i>orf63</i> from phage $\Phi24_B$	This study

GC) and  $\lambda orf63\_BamHI$  (5' GAG GAT CCT GCA TTC CGT GGT TGT C), and phage DNA as a template, which was isolated by using MasterPure™ Complete DNA and RNA Purification Kit (Epicenter). Then, the  $\lambda orf63$  was ligated with fragment of plasmid pUC18 (insert and vector were digested with EcoRI and BamHI restrictions endonucleases; Thermo Scientific), bearing an ampicillin resistance gene and sequence of *p<sub>lac</sub>* promoter. The plasmid pSB\_orf63\_ $\Phi24_B$  was constructed according to similar procedure. To amplify a DNA fragment containing *vb\_24B\_8c* sequence (the homolog of  $\lambda orf63$ ) by PCR method, two primers: F $\Phi24_B orf63\_EcoRI$  (5'GGA GAA TTC GGC TGT ATG CAC AAA GC) and R $\Phi24_B orf63\_BamHI$  (5'GTA GGA TCC TTG TCA TGC CGG GTC) were used. Next, plasmid pUC18 and insert were cut with EcoRI and BamHI enzymes and ligate by the T4 DNA ligase (Thermo Scientific). The construction of pUC18 derivatives was confirmed by DNA sequencing (Genomed).

### Media and Growth Conditions

All *in vivo* experiments were performed in LB liquid medium (Sigma—Aldrich) supplemented with 10 mM MgSO<sub>4</sub> (phage  $\lambda$  or phage  $\lambda\Delta orf63$ ) or with 10 mM MgSO<sub>4</sub> and 10 mM CaCl<sub>2</sub> (phage  $\Phi24_B$  or phage  $\Phi24_B\Delta orf63$ ), and with 50  $\mu$ g/ml ampicillin (if necessary) (Sigma—Aldrich). To stimulate Orf63 protein production from the recombinant pUC18 derivatives, overnight bacterial cultures were diluted 1:100 in fresh LB medium and treated with IPTG (A&A Biotechnology) to a final concentration of 1 mM. Then, host bacteria were grown in aeration condition, achieved by shaking, at 30°C to A<sub>600</sub> = 0.1 or 0.2 (the optical density of bacterial cells was dependent on the experimental conditions described in the following chapters).

### Double Overlay Plaque Assay

Bacteriophage titration was performed on the standard Petri dishes (Alchem) filled with 25 ml of LB agar (1.5% agar; Sigma—Aldrich), according to a procedure described by Sambrook and Russell (2001), with some modification. The top layer was prepared by mixing 2 ml of LB agar (0.7% agar; Sigma—Aldrich) with 1 ml of the overnight bacterial cell culture. To obtain visible plaques formed by Stx phages, the bottom agar was supplemented with sublethal concentration of chloramphenicol (Sigma—Aldrich). This antibiotic was effective in increasing of size of plaques of phage  $\Phi24_B$  and its derivative, which possessed in genomes chloramphenicol resistance gene (Table 1). As described previously (Loś et al., 2008a), the *cm* gene expression, especially after phage infection of *E. coli* bacteria, may have the positive influence on cellular productivity by decreasing the inhibitory effects of the antibiotic on protein synthesis. To determine the number of phages per ml of suspension (PFU/ml), serial 10-fold dilutions were prepared in TM buffer (10 mM Tris–HCl, 10 mM MgSO<sub>4</sub>; pH 7.2). Then, appropriate volume of each dilution of phage lysate was spotted onto double agar layer. The plates were incubated at 37°C overnight, plaques were counted, and the phage titer was calculated.

### One-Step Growth Experiments in Phage-Infected Bacteria

To investigate the intracellular lytic development of lambdoid phages the one-step-growth experiment was prepared using the method described by Wegrzyn et al. (1995), with a minor modification (Bloch et al., 2014; Nejman-Falenczyk et al., 2015). Host bacteria were grown in LB medium at 30°C to A<sub>600</sub> = 0.2. In

the next step, 10 ml of a bacterial culture was centrifuged ( $2,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ ). The pellet was suspended in 1 ml of LB medium supplemented with 3 mM sodium azide (Sigma—Aldrich). Bacteriophages were added to *E. coli* cells to m.o.i. of 0.05. After 10 min incubation at  $30^{\circ}\text{C}$ , unadsorbed phages were removed by three times washing in LB medium with 3 mM sodium azide ( $2,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ ). Then, 25  $\mu\text{l}$  of the suspension was added to 25 ml of LB medium prewarmed to  $30^{\circ}\text{C}$  (time 0) and cultivated in an incubator shaker. The number of infection centers were determined at times: 5, 10, 15 min after infection by mixing 0.1 ml of the sample with 0.9 ml of an overnight culture of appropriate indicator bacteria and 2 ml of top agar. Next, the mixture was poured onto LB agar plate (phages  $\lambda$  and  $\lambda\Delta\text{orf63}$ ) or LB agar plate with 2.5  $\mu\text{g/ml}$  chloramphenicol (phage  $\Phi 24_{\text{B}}$  and  $\Phi 24_{\text{B}}\Delta\text{orf63}$ ). Samples taken at later times were treated with chloroform (POCH), shaken vigorously and cleared by centrifugation ( $2,000 \times g$  for 5 min at RT). The phage lysate was diluted in TM buffer and titrated under permissive condition. Plates were incubated at  $37^{\circ}\text{C}$  overnight. The number of viruses released from each infected cell (burst size) was calculated as a ratio of phage titer to the titer of infection centers.

## Prophage Induction with Hydrogen Peroxide

Bacteria lysogenic for lambdoid phages were grown in LB medium at  $30^{\circ}\text{C}$  to  $A_{600} = 0.1$ . Next, the culture was divided into two aliquots. One of them was treated with 1 mM hydrogen peroxide (Sigma—Aldrich) to provoke the prophage induction. The second one was a control without an induction agent. The cultivation was continued at  $30^{\circ}\text{C}$ . At indicated times samples were harvested, mixed with chloroform and vortexed for 1 min. The suspension was centrifuged for 5 min in a microfuge at RT. The supernatant was diluted in TM buffer and 2.5  $\mu\text{l}$  of each serial dilution was dropped onto a freshly prepared double-layer LB agar in plastic Petri dishes. Plates were incubated at  $37^{\circ}\text{C}$  overnight. The relative phage titer was estimated by subtracting the phage titer determined in non-induced cultures from the phage titer estimated in induced cultures.

## Survival of Host Bacteria after Bacteriophage Infection

To estimate the percentage of surviving cells after bacteriophage infection the procedure created by Sambrook and Russell (2001) was used, with a minor modification. A bacterial culture was grown at  $30^{\circ}\text{C}$  to  $A_{600} = 0.2$ . Then, 4 ml of the sample was centrifuged ( $2,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ ). The supernatant was discarded and the pellet was washed with 0.85% sodium chloride (POCH) ( $2,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ ). Finally, the bacterial pellet was suspended in 1 ml of LB medium supplemented with  $\text{MgSO}_4$  (phage  $\lambda$  and phage  $\lambda\Delta\text{orf63}$ ) or with  $\text{MgSO}_4$  and  $\text{CaCl}_2$  (phage  $\Phi 24_{\text{B}}$  and  $\Phi 24_{\text{B}}\Delta\text{orf63}$ ) to a final concentration of 10 mM. The suspension was incubated for 30 min at  $30^{\circ}\text{C}$  and then phage particles were added to m.o.i. of 1, 5, 10. The mixture was kept for 15 min (phage  $\lambda$  and phage  $\lambda\Delta\text{orf63}$ ) or 30 min (phage  $\Phi 24_{\text{B}}$  and  $\Phi 24_{\text{B}}\Delta\text{orf63}$ ) at  $30^{\circ}\text{C}$ . In the next step, serial dilutions in

0.85% sodium chloride were prepared and 40  $\mu\text{l}$  of each dilution was spread on LB agar plates. After overnight incubation at  $37^{\circ}\text{C}$ , percentage of surviving *E. coli* bacteria was calculated relative to bacterial culture in which TM buffer was added instead of phage particles.

## Efficiency of Prophage Formation after Bacterial Virus Infection

Efficiency of lysogenization was estimated according to Arber et al. (1983) and Wegrzyn et al. (1992), with some modification. Host bacteria were cultured at  $30^{\circ}\text{C}$  to  $A_{600} = 0.2$ . Next, 1 ml of the sample was centrifuged ( $2,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ ). Bacterial culture was washed with TCM buffer twice, and then pellet was suspended in the same buffer. Bacteriophages were added to bacterial cells to m.o.i. of 1, 5, 10. The mixture was incubated at  $30^{\circ}\text{C}$ . Then, serial dilutions were prepared and 20  $\mu\text{l}$  of each suspension was spread on LB agar plates prior to overnight incubation at  $37^{\circ}\text{C}$ . The next day, 96 colonies were passaged in each well of a 96-well plate with 200  $\mu\text{l}$  of LB medium and shaken at  $37^{\circ}\text{C}$  to  $A_{600} = 0.1$ . To estimate a percent of lysogens among survivors, bacterial cultures were treated with UV light at  $50 \text{ J/m}^2$  (the dose used routinely for lambdoid prophage induction) and incubated at  $37^{\circ}\text{C}$  for 2 h. Following induction, putative lysogens were mixed with chloroform, centrifuged ( $2,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ ) and the water phase was spotted onto a double-layer LB agar (phage  $\lambda$  and phage  $\lambda\Delta\text{orf63}$ ) or a double-layer LB agar supplemented with chloramphenicol to a final concentration of 2.5  $\mu\text{g/ml}$  (phage  $\Phi 24_{\text{B}}$  and  $\Phi 24_{\text{B}}\Delta\text{orf63}$ ). Efficiency of lysogenization was calculated as a percent of lysogens relative to all tested bacterial cells. Lysogens were also infected with the same phage to check their resistance to superinfection, as described previously (Wegrzyn et al., 1992).

## Prophage Induction and Extraction of RNA

Induction of tested prophages was provoked in lysogenic bacteria by addition of hydrogen peroxide to a final concentration of 1 mM. At the appropriate time,  $10^9$  bacterial cells were harvested, treated with 10 mM sodium azide and deep frozen in liquid nitrogen (this procedure was necessary to inhibit the growth of host bacteria). Total RNA from all samples were isolated with the High Pure RNA Isolation Kit (Roche Applied Science). To remove DNA from RNA preparations the TURBO DNA-free™ Kit (Life Technologies) was used. The quality and quantity of total isolated RNA were analyzed by a NanoDrop spectrophotometer and agarose gel electrophoresis. The contamination of DNA from RNA samples was also tested by routine PCR and qRT-PCR.

## cDNA Synthesis from an RNA Template

To synthesize cDNA from an RNA template, the Transcriptor Reverse Transcriptase and random hexamer primers (Roche Applied Science) were used, according to the protocol supplied from the provider. 1.25  $\mu\text{g}$  of the total RNA was taken for each reaction. Finally, mixture was diluted 10-fold and tested in qRT-PCR.

**TABLE 2 |** Primers used for RT-Qpcr.

Primer name	Sequence (5' → 3')
pF_λ_ea8.5	GGGCAAGTATCGTTTCCACC
pR_λ_ea8.5	GCAATGTGCGAGAAATGACTG
pF_λ_ea22	GCAGTTCCAGCACAATCGAT
pR_λ_ea22	AATGCATGACGACTGGGGAT
pF_λ_orf73	CACTTCGAACCTCTCTGTTTACTG
pR_λ_orf73	CAGGGTTGTGCGGACTTGTG
pF_λ_orf61	TTAGCCTGACGGGCAATG
pR_λ_orf61	CCGACATGGGACTTGTTC
pF_λ_orf60a	GCATACAGCCCCCTCGTTTAT
pR_λ_orf60a	CCGAAATCCACTGAAAGCAC
pF_λ_cIII	ATTCTTTGGGACTCCTGGCTG
pR_λ_cIII	GTAAATTACGTGACGGATGAAAC
pF_λ_N	CTCGTGATTTTCGGTTTGCGA
pR_λ_N	AAGCAGCAAATCCCCTGTTG
pF_λ_cro	ATGCGGAAGAGGTAAGCCCC
pR_λ_cro	TGGAATGTGTAAGAGCGGGG
pF_λ_cII	TCGCAATGCTTGGAAGTGA
pR_λ_cII	CCCTCTCCACCTGCTGATC
pF_λ_Q	TTCTGCGGTAAGCACGAAC
pR_λ_Q	TGCATCAGATAGTTGATAGCCTT
pF_λ_R	ATCGACCGTTGACGCAATA
pR_λ_R	GCTCGAACTGACCATAACCAG
pF_Φ24B_ea22	TCAGCAACATGGCATTCACT
pR_Φ24B_ea22	GGTTGGGAAGCTGAGAGTTG
pF_Φ24B_orf73	CGAACCTCTCTGTTTACTGATAAGC
pR_Φ24B_orf73	TTACAGGTTGTGCGACTTGT
pF_Φ24B_orf61	TTAGCCTGACGGGCAATG
pR_Φ24B_orf61	CCGACATGGGACTTGTTC
pF_Φ24B_orf60a	CATACAGCCCCCTCGTTTAT
pR_Φ24B_orf60a	CCGAAATCCACTGAAAGCAC
pF_Φ24B_cIII	ATTCTTTGGGACTCCTGGCTG
pR_Φ24B_cIII	GTAAATTACGTGACGGATGAAAC
pF_Φ24B_N	AGGCGTTTCGTGAGTACCTT
pR_Φ24B_N	TTACACCGCCCTACTCTAAGC
pF_Φ24B_cro	CGAAGGCTTGTGGAGTTAGC
pR_Φ24B_cro	GTCTTAGGGAGGAAGCCGTT
pF_Φ24B_cII	TGATCGCGCAGAACTGATTTAC
pR_Φ24B_cII	GACAGCCAATCATCTTTGCCA
pF_Φ24B_O	AAGCGAGTTTGCCACGAT
pR_Φ24B_O	GAACCCGAAGTCTTACCG
pF_Φ24B_Q	GGGAGTGAGGCTTGAGATGG
pR_Φ24B_Q	TACAGAGGTTCTCCCTCCCG
pF_Φ24B_R	GGGTGGATGGTAAGCCTGT
pR_Φ24B_R	TAACCCGGTCGCAATTTTC
pF_E.coli_16SrRNA	CCTTACGACAGGGCTACAC
pR_E.coli_16SrRNA	TTATGAGGTCCGCTTGCTCTC

## qRT-PCR Assay and Data Analysis

The pattern of genes expression after prophage induction was performed by using the LightCycler® 480 Real-Time PCR System (Roche Applied Science), LightCycler® 480 SYBR Green I Master (Roche Applied Science) and cDNA samples.

Transcription rates of genes of lambdoid bacteriophages were compared in parallel to the 16S rRNA housekeeping gene (according to a procedure described by Strauch et al. (2008), which expression was stable during prophage induction provoked by hydrogen peroxide. All primers were created by Primer3web version 4.0.0 and are listed in **Table 2**. Each reaction mixture consisted of: 2x SYBR Green I Master Mix, 6.25 ng/μl cDNA and 200 nM specific primers. qRT-PCR amplifications were performed for 55 cycles. To confirm the specificity of primers, melting curve for each product was analyzed. The relative changes in gene expressions were determined by E-Method and calculated by the following formula: Normalized relative ratio =  $E_t^{CT(\text{target}) \text{ calibrator} - CT(\text{target}) \text{ sample}} / E_r^{CT(\text{reference}) \text{ calibrator} - CT(\text{reference}) \text{ sample}}$ , where  $E_t$  is the PCR efficiency of target and  $E_r$  means the PCR efficiency of reference. The sample before the addition of the inductor (the time point “zero”) was a calibrator. The raw run data for tested lambdoid phages were transferred using the “LC480 Conversion: conversion of raw LC480 data” software and then, PCR efficiency for each gene was calculated by LinRegPCR program, which was successfully used previously (Bloch et al., 2014, 2015; Nejman-Falenczyk et al., 2015; Licznarska et al., 2016a).

## Statistical Analysis

Each experiment was repeated three times and variation among replicates was presented as the error bars indicating the standard deviation (SD). All data comparisons were made by using Student's *t*-test. Significant differences were marked by asterisks when  $P < 0.05$  (\*) or  $P < 0.01$  (\*\*).

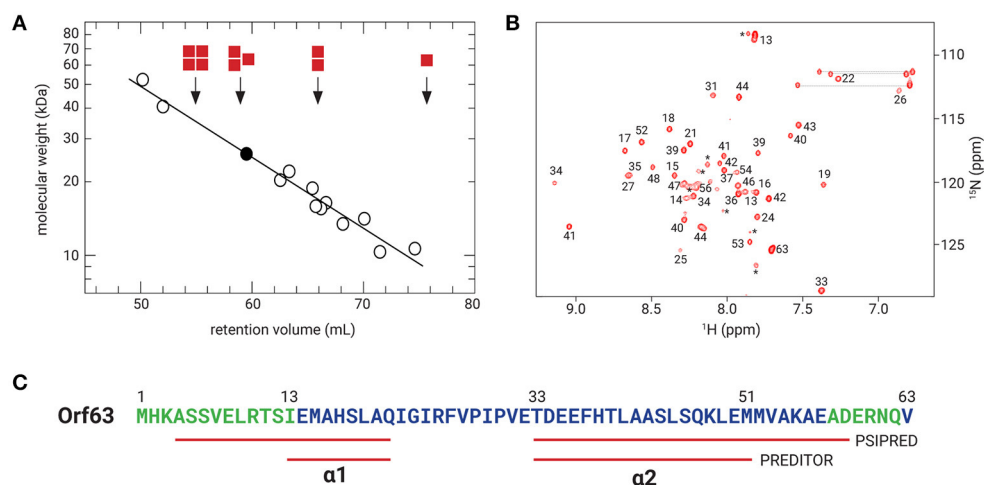
## RESULTS

### The Oligomeric State of Orf63

Samples from four independent preparations of 6xHis-Flag tagged Orf63 (15 aa. tag + 63 aa. protein) eluted as one peak on a preparative gel filtration column with an average retention volume of 59.5 mL corresponding to an apparent molecular weight of ~26 kDa (**Figure 2A**). Since affinity-tagged Orf63 is only 9 kDa, the gel filtration results suggest that Orf63 is oligomeric with a trimer as the most plausible configuration. This estimate is most accurate if Orf63 has the characteristics of a globular protein to match the standards used.

### The orf63 Gene Encodes a Folded Protein

Consistent with the observation that Orf63 is oligomeric in solution, NMR spectra of Orf63 at 298 K (25°C) suffered from considerable resonance line broadening that was characteristic of proteins > 20 kDa in overall molecular weight. Consequently, a higher temperature of 310 K (37°C) was chosen for all NMR studies to increase the tumbling time of the protein that, in turn, improves the sensitivity of triple resonance experiments. In **Figure 2B**, a  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum is presented. The amide resonances in this two-dimensional spectrum are disperse indicating that the protein is folded. The combined analysis of several triple resonance ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) spectra lead to the determination of backbone (HN,



**FIGURE 2 |** Orf63 structural insights. **(A)** The retention volumes for a set of proteins ranging from 10–67 kDa were plotted against molecular weight to produce the log-linear relationship shown (open circles). The  $K_{av}$  value observed for Orf63 corresponds to an apparent molecular weight of ~28 kDa (closed circle). For reference, red squares denote the expected retention volumes of monomeric, dimeric, trimeric, and tetrameric species of the 9 kDa affinity tagged Orf63 protein. **(B)** A representative  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Orf63 at 37°C demonstrating the characteristic resonance dispersion of a folded protein. Unassigned residues are labeled with an asterisk. Lines denote the amino resonances of three asparagine and glutamine side chains. **(C)** A conventional strategy was used to assign the backbone (HN, N, CA, CB, C') chemical shifts of Orf63. The extent of these assignments is colored blue on the sequence. The chemical shift data were then used as input to PREDITOR (Berjanskii et al., 2006) which identified two helices labeled  $\alpha 1$  and  $\alpha 2$ . Using sequence data alone as input to PSIPRED (McGuffin et al., 2000), longer helices are predicted.

N, CA, CB, C') chemical shift assignments for residues 14–52 of Orf63. Resonances for the amino terminal affinity tags and from residue 53 onwards to the carboxy-terminus were either not observed or unassignable. Thus, the NMR data suggest that the folded region of Orf63 includes from residues 14–52.

## Structural Characteristics of Orf63

Several statistical methods are available to predict secondary structure from backbone chemical shift data with a high degree of accuracy. As shown in **Figure 2C**, two helices are predicted ( $\alpha 1$ : 13–21;  $\alpha 2$ : 33–50). The secondary structure determined from chemical shift data is consistent with the secondary structure of Orf63 predicted from sequence information alone, although the helical boundaries are different.

## The Sequence of Putative *orf63* Products is Conserved among Lambdoid Bacteriophages

Since experiments shown in **Figure 2** indicated that *orf63* encodes a protein, we have tested similarity of the putative proteins encoded by *orf63* of different lambdoid bacteriophages. Thus, scores of pairwise alignments of the predicted amino acid sequences of *orf63* from six such phages have been calculated. As demonstrated in **Table 3**, all these putative proteins are similar to each other. This indicates that the high similarity is kept at the protein level of Orf63 of lambdoid bacteriophages.

**TABLE 3 |** Scores of pairwise alignments of the predicted amino acid sequences of *orf63* from six analyzed lambdoid phages:  $\lambda$  phage (NC\_001416),  $\Phi 24_B$  phage (HM208303), 933W phage (NC\_000924), VT2 Sakai phage (AP000422), Stx1 converting phage (NC\_004913), and Stx2 converting phage II (NC\_004914).

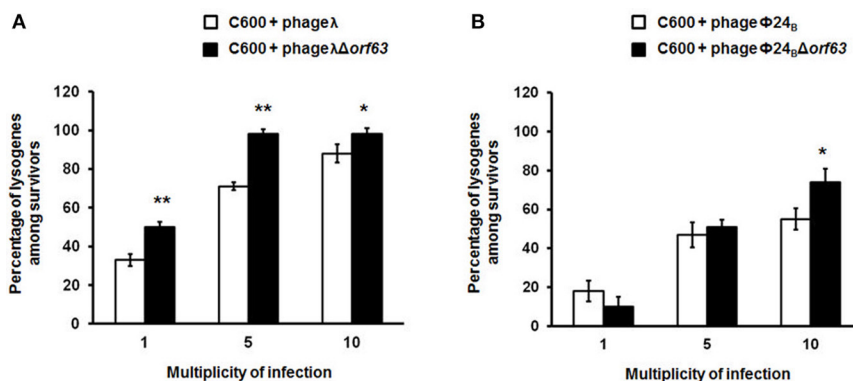
	$\lambda$	$\Phi 24_B$	933W	VT2 Sakai	Stx1	Stx2_II
$\lambda$		86	86	87	87	87
$\Phi 24_B$			100	98	98	98
933W				98	98	98
VT2 Sakai					100	100
Stx1						100
Stx2_II						

Pairwise scores are simply the number of identities between the two sequences, divided by the length of the alignment, and represented as a percentage. The multiple sequence alignment was performed using the ClustalW algorithm.

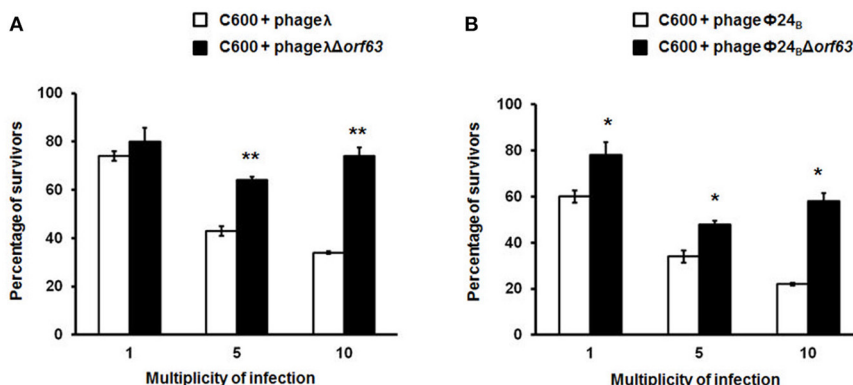
## Efficiency of Lysogenization and Prophage Induction in the Absence of *orf63*

Since previous studies suggested that genes from the *exo-xis* region might be involved in the regulation of bacteriophage development (Bloch et al., 2013, 2014; Licznarska et al., 2016a), we have tested two crucial controlled steps in the lambdoid phage life cycle, the lysis-vs.-lysogenization decision, and prophage induction. We found that lysogenization efficiency was significantly increased in bacteriophages  $\lambda$  and  $\Phi 24_B$  devoid of *orf63* (**Figures 3A,B**, respectively) though this phenomenon was more pronounced in  $\lambda$  (the effects were seen at all tested m.o.i.) (**Figure 3A**) than in  $\Phi 24_B$  (significant effects observed only at m.o.i. = 10) (**Figure 3B**). Also, survival rates of bacterial cells (i.e., cells lysogenized and not infected) in populations

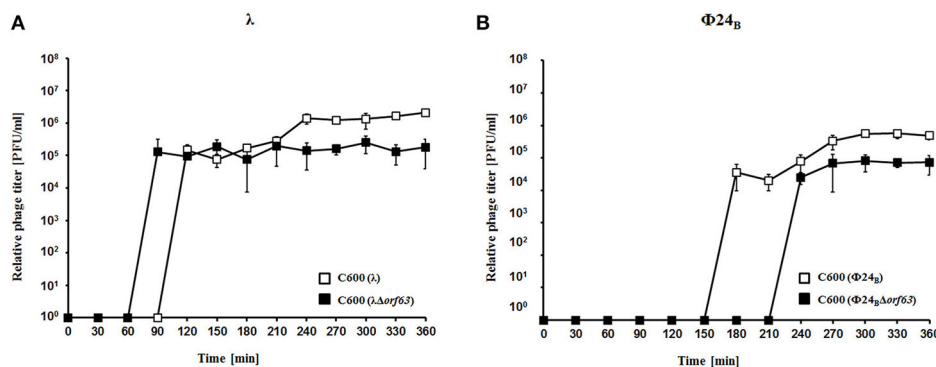




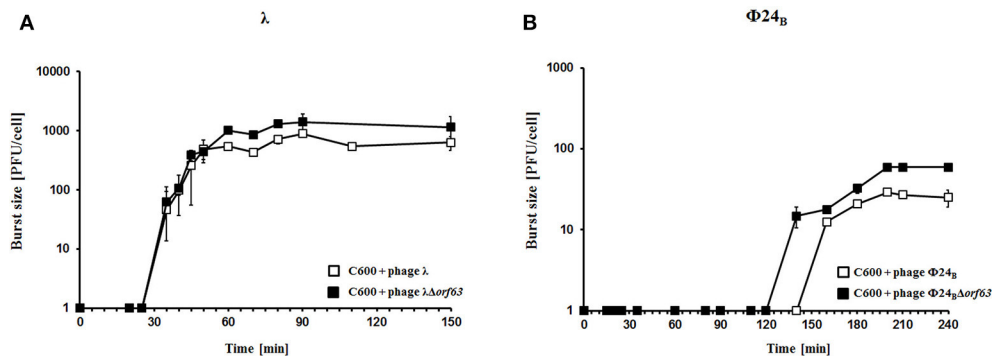
**FIGURE 3 |** Efficiency of lysogenization of *E. coli* C600 strain with lambdoid bacteriophages:  $\lambda$  and  $\Phi 24_B$  ( $\square$  in (A,B), respectively) or their deletion mutants  $\lambda\Delta orf63$  and  $\Phi 24_B\Delta orf63$  ( $\blacksquare$  in (A,B), respectively). Results are presented as mean values  $\pm SD$  from three independent experiments. Statistical analysis (*t* test) was performed for results from each m.o.i. (multiplicity of infection) between wild type phage and its deletion mutant. Significant differences are marked by asterisks  $P < 0.05$  (\*) or  $P < 0.01$  (\*\*).



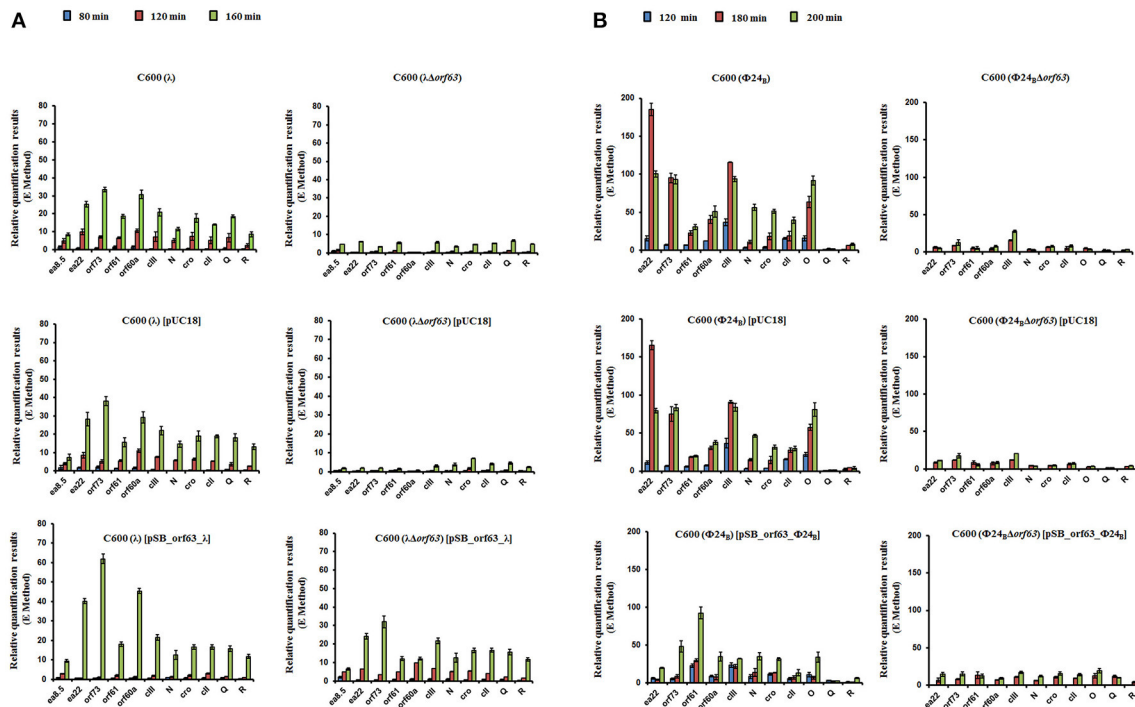
**FIGURE 4 |** Survival (%) of the wild-type strain *E. coli* C600, after infection with lambdoid bacteriophages:  $\lambda$  and  $\Phi 24_B$  ( $\square$  in (A,B), respectively) or their deletion mutants  $\lambda\Delta orf63$  and  $\Phi 24_B\Delta orf63$  ( $\blacksquare$  in (A,B), respectively). Mean values from three independent experiments  $\pm SD$  are shown. Statistical analysis were performed for each m.o.i. by *t* test. The significance of differences between fractions of bacterial cells surviving the infection with  $\lambda$  and  $\lambda\Delta orf63$  as well as  $\Phi 24_B$  and  $\Phi 24_B\Delta orf63$  are observed and marked by asterisks  $P < 0.05$  (\*) or  $P < 0.01$  (\*\*).



**FIGURE 5 |** Development of  $\lambda$  and  $\lambda\Delta orf63$  (A) or  $\Phi 24_B$  and  $\Phi 24_B\Delta orf63$  (B) bacteriophages after prophage induction with hydrogen peroxide at 30 °C. *E. coli* C600 bacteria were either lysogenic with wild-type phages  $\lambda$  and  $\Phi 24_B$  ( $\square$  in (A,B), respectively) or their deletion mutants  $\lambda\Delta orf63$  and  $\Phi 24_B\Delta orf63$  ( $\blacksquare$  in (A,B), respectively). Phage lytic development was initiated by addition of  $H_2O_2$  to final concentration of 1 mM at time 0. The presented results are mean values from three independent experiments with error bars indicating *SD* (note that in the most cases, the bars are smaller than sizes of symbols). Results are shown as PFU (plaque forming units) per one ml.



**FIGURE 6 |** Development of  $\lambda$  and  $\lambda\Delta orf63$  (A) or  $\Phi 24_B$  and  $\Phi 24_B\Delta orf63$  (B) bacteriophages following phage infection of *E. coli* bacteria. Host *E. coli* strains were infected with wild-type phages  $\lambda$  and  $\Phi 24_B$  ( $\square$  in A,B, respectively) or their deletion mutants  $\lambda\Delta orf63$  ( $\blacksquare$  in A,B, respectively) at time 0. The presented results are mean values  $\pm$ SD from three independent experiments. Results are shown as PFU (plaque forming units) per cell.

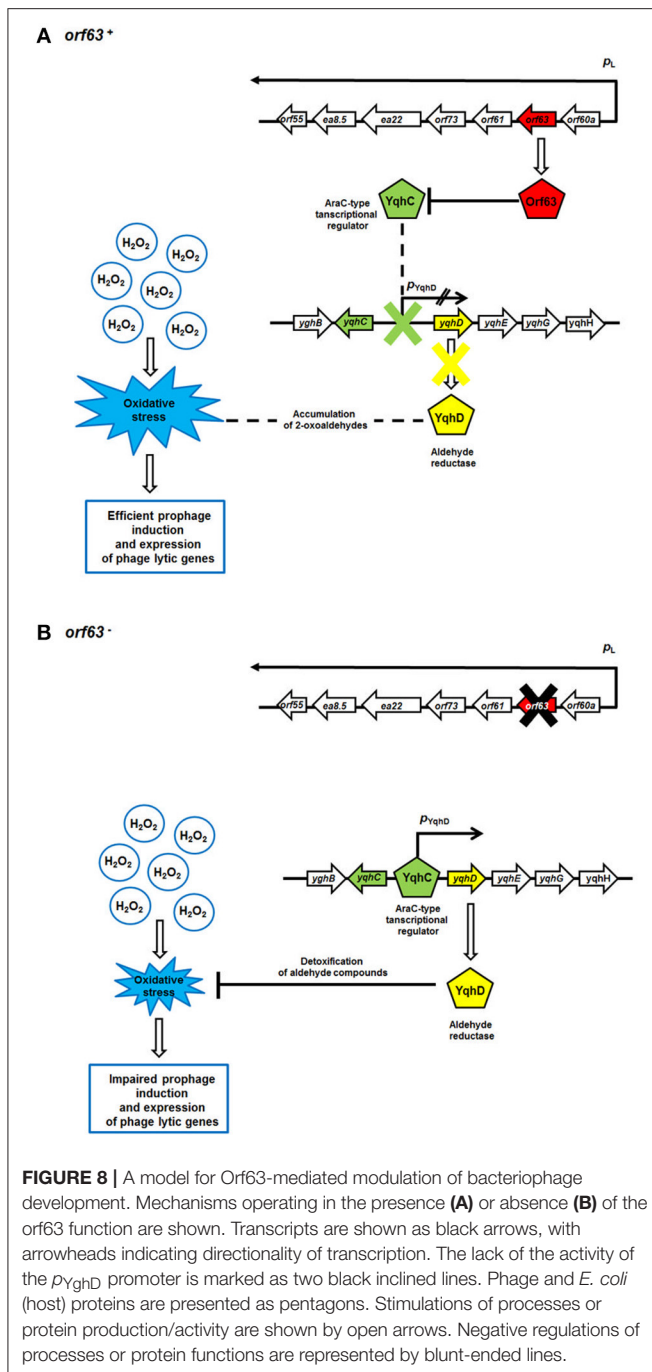


**FIGURE 7 |** Levels of transcripts of indicated genes of wild type phages  $\lambda$  and  $\Phi 24_B$  and their deletion mutants in the absence or presence of control plasmid [pUC18] or plasmid bearing the deleted *orf63* [pSB\_orf63], assessed by quantitative reverse RT-PCR analysis, after prophage induction with 1 mM  $H_2O_2$  in *E. coli* C600 host at 30°C. Levels of transcripts corresponding to particular genes were determined at following times after induction: 80 (blue), 120 (red) or 160 (green) minutes in case of phages  $\lambda$  and  $\lambda\Delta orf63$  (A), and 120 (blue), 180 (red) or 200 (green) minutes in case of phages  $\Phi 24_B$  and  $\Phi 24_B\Delta orf63$  (B). Results are presented as mean values from three independent experiments with error bars indicating SD.

infected with  $\Delta orf63$  or  $\Phi 24_B \Delta orf63$  were higher than those in experiments with wild-type  $\lambda$  or  $\Phi 24_B$  (Figures 4A,B, respectively), supporting the conclusion that lysogenization is more effective for the mutant, indeed.

To test efficiency of prophage induction, we have estimated the number of phages appearing after prophage induction with hydrogen peroxide (one of natural prophage inducers occurring in human intestine, the common habitat of *E. coli*). Deletion of *orf63* caused a lower phage titer after prophage induction for

both  $\lambda$  and  $\Phi 24_B$  (Figures 5A,B, respectively). However, when measured kinetics of phage development following infection of *E. coli* cells at low m.o.i. (0.05), we found that phages  $\lambda$  and  $\Phi 24_B$  devoid of *orf63* gave even more progeny per infected cell than their wild-type counterparts (Figures 6A,B, respectively). Therefore, combining results of experiments presented in Figures 5, 6, one can conclude that deletion of *orf63* influences efficiency of prophage induction in both  $\lambda$  and  $\Phi 24_B$ . Since formation of progeny phages is definitely not impaired in the



absence of *orf63* when lytic development starts after infection, we suggest that lower phage titer after prophage induction indicates lower efficiency of this process in phages devoid of this gene.

## Deletion of *Orf63* Influences Expression of Genes of $\lambda$ and $\Phi 24_B$ Phages

Since experiments described above indicated that *orf63* function is involved in the regulation of lysogenization and prophage induction in both  $\lambda$  and  $\Phi 24_B$ , we aimed to measure

expression of selected bacteriophage genes in *E. coli* cells after hydrogen peroxide-provoked prophage induction. Reverse transcription quantitative real time PCR (RT-qPCR) was used to assess abundance of particular transcripts. We have measured expression levels of genes from the *exo-xis* region (*ea8.5*, *ea22*, *orf73*, *orf61*, *orf60a*) and some key regulatory genes of  $\lambda$  and  $\Phi 24_B$ , i.e., *N*, *cro*, *cII*, *Q*, *R*. We found that expression of all tested genes was significantly impaired in  $\Delta orf63$  mutants of both  $\lambda$  and  $\Phi 24_B$  relative to wild-type phages at all tested times after prophage induction (Figures 7A,B, respectively). These results confirm that prophage induction is significantly impaired in the absence of *orf63*, and suggest a regulatory role for the *orf63* gene product in the control of expression of phage genes. In the case of phage  $\lambda$ , complementation of the  $\Delta orf63$  mutation by overexpression of wild-type *orf63* from a plasmid was successful, at least at certain times after prophage induction (Figure 7A). However, we failed to obtain such a complementation in phage  $\Phi 24_B$  (Figure 7B). This might suggest that specific ratio(s) of Orf63 is/are required for accurate regulation of phage development.

## DISCUSSION

Considering that the complete 48,502 bp genome of  $\lambda$  was achieved in 1983, it is astounding that this well-investigated model virus still contains uncharacterized open reading frames, many of which lie between the *exo*, and *xis* genes. We began this investigation by demonstrating that *orf63* found within the *exo-xis* region, encodes a *bona fide* protein both in structural, and functional terms. Consistent with its name, this small 63 aa. protein is comprised of only two helices that cover most of the available sequence. Analytical gel filtration of purified Orf63 suggests that the oligomeric state is a trimer. If Orf63 deviates significantly from a globular shape, it is possible that the molecular weight may be overestimated by the gel filtration assay, the oligomeric state could be a dimer. However, given the significant line broadening observed during a series of initial NMR based surveys performed at 25°C that could only be alleviated by performing all of the studies subsequently at 37°C, the NMR data tend to corroborate the gel filtration findings.

Since Orf63 appears to be a functional protein, we tested effects of deletion of *orf63* on development of bacteriophages  $\lambda$  and  $\Phi 24_B$ . In the absence of functional Orf63, we observed a significant increase in the efficiency of lysogenization, and considerable lower efficiency of hydrogen peroxide-mediated prophage induction. These results may suggest that Orf63 is involved in the regulation of expression of specific phage genes. Studies with the use of RT-qPCR revealed that expression of vast majority of crucial regulatory genes, as well as genes from the *exo-xis* region, is significantly influenced by the absence of *orf63*. Moreover, perhaps specific ratio of Orf63 to other regulators is required, as it was impossible to obtain complementation with the wild-type *orf63* expressed from a plasmid in  $\Phi 24_B$ , though it was successful in  $\lambda$ . The hypothesis about the requirement of specific Orf63 ratio(s) for accurate regulation of phage

development is supported by impaired expression of  $\Phi 24_B$  genes during overexpression of *orf63*.

A previous yeast two-hybrid study identified the transcription factor YqhC as a possible protein partner of Orf63 (Blasche et al., 2013). YqhC is interesting because it is a transcription factor that promotes the synthesis of YqhD, the major enzyme responsible for detoxifying compounds produced from glucose under conditions of oxidative stress (Lee et al., 2010). Among the compounds that YqhD acts upon are 2-oxoaldehydes, toxic and highly reactive products of oxidative stress on the bacterium formed from glucose. It has been proposed that oxoaldehydes are one class of compound that is capable of inducing a wider stress response through SoxRS (Benov and Fridovich, 2002). Since neutrophils mount a vigorous oxidative attack during a STEC infection, Orf63 may be beneficial to the bacteriophage by manipulating the microenvironment of the bacterial host. Orf63 by binding the transcriptional activator YqhC, may prevent it from attenuating the stress response created by neutrophil mediated attack on EHEC strains in the gut, and promoting a transition to the lytic phase commensurate with the activation of associated phage genes in that response. A putative mechanism for Orf63-mediated modulation of prophage induction and phage genes' expression is presented in **Figure 8**. Notwithstanding its role in protein-protein interactions, it is still possible that Orf63 itself could function as a transcriptional regulator, although its small size, lack of a known DNA binding domain, and relatively few basic amino acids argue against this possibility. A high-resolution structure of Orf63 alone, or in complex with a possible interactor like YqhC, will resolve these outstanding questions.

## REFERENCES

- Allison, H. E. (2003). Immunity profiles of wild-type and recombinant Shiga-like toxin-encoding bacteriophages and characterization of novel double lysogens. *Infect. Immun.* 71, 3409–3418. doi: 10.1128/IAI.71.6.3409-3418.2003
- Appleyard, R. K. (1954). Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. *Genetics* 39, 440–452.
- Arber, W., Enquist, L., Hohn, B., Murray, N. E., and Murray, K. (1983). "Lambda II," in *Experimental Methods for Use with  $\lambda$* , eds J. W. Roberts, F. W. Stahl, and R. A. Weisberg (New York, NY: Cold Spring Harbor Laboratory Press), 433–466.
- Benov, L., and Fridovich, I. (2002). Induction of the soxRS regulon of *Escherichia coli* by glycolaldehyde. *Arch. Biochem. Biophys.* 407, 45–48. doi: 10.1016/S0003-9861(02)00498-8
- Berjanskii, M. V., Neal, S., and Wishart, D. S. (2006). PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res.* 34, W63–W69. doi: 10.1093/nar/gkl341
- Blasche, S., Wuchty, S., Rajagopala, S. V., and Uetz, P. (2013). The protein interaction network of bacteriophage lambda with its host, *Escherichia coli*. *J. Virol.* 87, 12745–12755. doi: 10.1128/JVI.02495-13
- Bloch, S. K., Felczykowska, A., and Nejman-Falencyk, B. (2012). *Escherichia coli* O104:H4 out break- have we learnt a lesson from it? *Acta Biochim. Pol.* 59, 483–488.
- Bloch, S., Nejman-Falencyk, J., Loś, J. M., Baranska, S., Lepek, K., Felczykowska, A., et al. (2013). Genes from the exo-xis region of  $\lambda$  and Shiga toxin-converting bacteriophages influence lysogenization and prophage induction. *Arch. Microbiol.* 195, 693–703. doi: 10.1007/s00203-013-0920-8
- Bloch, S., Nejman-Falencyk, B., Dydecka, A., Loś, J. M., Felczykowska, A., Węgrzyn, A., et al. (2014). Different expression patterns of genes from the exo-xis region of bacteriophage  $\lambda$  and Shiga toxin-converting bacteriophage  $\Phi 24B$  following infection or prophage induction in *Escherichia coli*. *PLoS ONE* 9:e108233. doi: 10.1371/journal.pone.0108233
- Bloch, S., Nejman-Falencyk, B., Topka, G., Dydecka, A., Licznarska, K., Narajczyk, M., et al. (2015). UV-Sensitivity of Shiga Toxin-Converting Bacteriophage Virions  $\Phi 24B$ , 933W, P22, P27 and P32. *Toxins* 7, 3727–3739. doi: 10.3390/toxins7093727
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293. doi: 10.1007/BF00197809
- Hendrix, R. W., and Duda, R. L. (1992). Bacteriophage lambda PaPa: not a mother of all lambda phages. *Science* 258, 1145–1148. doi: 10.1126/science.1439823
- Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H., and Wagner, G. (2012). Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR* 52, 315–327. doi: 10.1007/s10858-012-9611-z
- Jensen, K. F. (1993). The *Escherichia coli* K-12 wild types W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *J. Bact.* 175, 3401–3407. doi: 10.1128/jb.175.11.3401-3407.1993
- Kwan, J. J., Smirnova, E., Khazai, S., Evanics, F., Maxwell, K. L., and Donaldson, L. W. (2013). The solution structures of two prophage homologues of the bacteriophage  $\lambda$  Ea8.5 protein reveal a newly discovered hybrid homeodomain/zinc-finger fold. *Biochemistry* 52, 3612–3614. doi: 10.1021/bi400543w
- Lee, C., Kim, I., Lee, J., Lee, K. L., Min, B., and Park, C. (2010). Transcriptional activation of the aldehyde reductase YqhD by YqhC and its implication in glyoxal metabolism of *Escherichia coli* K-12. *J. Bacteriol.* 192, 4205–4214. doi: 10.1128/JB.01127-09
- Licznarska, K., Dydecka, A., Bloch, S., Topka, G., Nejman-Falencyk, B., Węgrzyn, A., et al. (2016a). The Role of the exo-xis region in oxidative stress-mediated

## AUTHOR CONTRIBUTIONS

AD and SB contributed equally to this work. AD, SB, BN, GT, TG, and AN took part in the physiological studies on bacteria and phages and data processing. AD, SB, and BN participated also in writing the manuscript and planning of the study. AR, SP, and LD performed the protein analyses. GW, LD, and AW were engaged in writing the manuscript and took part in planning of the study and discussions.

## ACKNOWLEDGMENTS

This work was supported by the National Science Center (Poland), grants no. UMO-2013/09/B/NZ2/02366 and UMO-2015/17/B/NZ9/01724 to AW, and by the Faculty of Biology, University of Gdansk, grant no. 538-L140-B539-17 to AD. The Natural Sciences and Engineering Research Council (NSERC) is acknowledged for support to LD (Discovery Operating Grant 238924), and AR/SP (USRA training awards).



- induction of Shiga toxin-converting prophages. *Oxid. Med. Cell. Longev.* 2016:8453135. doi: 10.1155/2016/8453135
- Licznarska, K., Nejman-Falenczyk, B., Bloch, S., Dydecka, A., Topka, G., Gasior, T., et al. (2016b). Oxidative stress in Shiga toxin production by enterohemorrhagic *Escherichia coli*. *Oxid. Med. Cell. Longev.* 2016:3578368. doi: 10.1155/2016/3578368
- Łoś, J. M., Golec, P., Węgrzyn, G., Węgrzyn, A., and Łoś, M. (2008a). Simple method for plating *Escherichia coli* bacteriophages forming very small plaques or no plaques under standard conditions. *Appl. Environ. Microbiol.* 74, 5113–5120. doi: 10.1128/AEM.00306-08
- Łoś, J. M., Łoś, M., Węgrzyn, A., and Węgrzyn, G. (2008b). Role of the bacteriophage  $\lambda$  exo-xis region in the virus development. *Folia Microbiol.* 53, 443–450. doi: 10.1007/s12223-008-0068-0
- McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405. doi: 10.1093/bioinformatics/16.4.404
- Mizutani, S., Nakazono, N., and Sugino, Y. (1999). The so-called chromosomal verotoxin genes are actually carried by defective prophages. *DNA Res.* 6, 141–143. doi: 10.1093/dnares/6.2.141
- Muniesa, M., Hammerl, J. A., Hertwig, S., Appel, B., and Brüssow, H. (2012). Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Appl. Environ. Microbiol.* 78, 4065–4073. doi: 10.1128/AEM.00217-12
- Nejman-Falenczyk, B., Bloch, S., Licznarska, K., Dydecka, A., Felczykowska, A., Topka, G., et al. (2015). A small, microRNA-size, ribonucleic acid regulating gene expression and development of Shiga toxin-converting bacteriophage  $\Phi$ 24<sub>B</sub>. *Sci. Rep.* 5:10080. doi: 10.1038/srep1008
- Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.
- Riley, L. W., Remis, R. S., Helgerson, S. D., McGee, H. B., Wells, J. G., and Davis, B. R. (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* 308, 681–685. doi: 10.1056/NEJM198303243081203
- Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd Edn. New York, NY: Cold Spring Harbor Laboratory Press.
- Schmidt, H. (2001). Shiga-toxin-converting bacteriophages. *Res. Microbiol.* 152, 687–695. doi: 10.1016/S0923-2508(01)01249-9
- Skinner, S. P., Goult, B. T., Fogh, R. H., Boucher, W., Stevens, T. J., Laue, E. D., et al. (2015). Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. D Biol. Crystallogr.* 71, 154–161. doi: 10.1107/S1399004714026662
- Strauch, E., Hammerl, J. A., Konietzny, A., Schneiker-Bekel, S., Arnold, W., et al. (2008). Bacteriophage 2851 is a prototype phage for dissemination of the Shiga toxin variant gene 2c in *Escherichia coli* O157:H7. *Infect. Immun.* 76, 5466–5477. doi: 10.1128/IAI.00875-08
- Węgrzyn, G., Glass, R. E., and Thomas, M. S. (1992). Involvement of the *Escherichia coli* RNA polymerase  $\alpha$  subunit in transcriptional activation by the bacteriophage  $\lambda$  CI and CII proteins. *Gene* 122, 1–7. doi: 10.1016/0378-1119(92)90025-K
- Węgrzyn, G., Węgrzyn, A., Konieczny, I., Bielawski, K., Konopa, G., Obuchowski, M., et al. (1995). Involvement of the host initiator function dnaA in the replication of coliphage lambda. *Genetics* 139, 1469–1481.
- Węgrzyn, G., Licznarska, K., and Węgrzyn, A. (2012). Phage  $\lambda$ —new insights into regulatory circuits. *Adv. Virus. Res.* 82, 155–178. doi: 10.1016/B978-0-12-394621-8.00016-9

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RO and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Dydecka, Bloch, Rizvi, Perez, Nejman-Falenczyk, Topka, Gasior, Necel, Węgrzyn, Donaldson and Węgrzyn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Transcriptome of *Streptococcus pneumoniae* Induced by Local and Global Changes in Supercoiling

Adela G. de la Campa<sup>1,2\*</sup>, María J. Ferrándiz<sup>1</sup>, Antonio J. Martín-Galiano<sup>1</sup>,  
María T. García<sup>3</sup> and Jose M. Tirado-Vélez<sup>1</sup>

<sup>1</sup> Unidad de Genética Bacteriana, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain,

<sup>2</sup> Presidencia, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>3</sup> Departamento de Microbiología, Facultad de Ciencias Biológicas, Universidad Complutense, Madrid, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Molecular Microbiology and Infection  
Biology (CIB, CSIC), Spain

### Reviewed by:

Andrea M. Mitchell,  
University of Birmingham,  
United Kingdom  
Jorge Bernardo Schwartzman,  
Cellular and Molecular Biology  
(CIB, CSIC), Spain

### \*Correspondence:

Adela G. de la Campa  
agcampa@isciii.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 24 May 2017

**Accepted:** 17 July 2017

**Published:** 31 July 2017

### Citation:

de la Campa AG, Ferrándiz MJ,  
Martín-Galiano AJ, García MT and  
Tirado-Vélez JM (2017)  
The Transcriptome of *Streptococcus*  
*pneumoniae* Induced by Local  
and Global Changes in Supercoiling.  
Front. Microbiol. 8:1447.  
doi: 10.3389/fmicb.2017.01447

The bacterial chromosome is compacted in a manner optimal for DNA transactions to occur. The degree of compaction results from the level of DNA-supercoiling and the presence of nucleoid-binding proteins. DNA-supercoiling is homeostatically maintained by the opposing activities of relaxing DNA topoisomerases and negative supercoil-inducing DNA gyrase. DNA-supercoiling acts as a general *cis* regulator of transcription, which can be superimposed upon other types of more specific *trans* regulatory mechanism. Transcriptomic studies on the human pathogen *Streptococcus pneumoniae*, which has a relatively small genome (~2 Mb) and few nucleoid-binding proteins, have been performed under conditions of local and global changes in supercoiling. The response to local changes induced by fluoroquinolone antibiotics, which target DNA gyrase subunit A and/or topoisomerase IV, involves an increase in oxygen radicals which reduces cell viability, while the induction of global supercoiling changes by novobiocin (a DNA gyrase subunit B inhibitor), or by seconeolitsine (a topoisomerase I inhibitor), has revealed the existence of topological domains that specifically respond to such changes. The control of DNA-supercoiling in *S. pneumoniae* occurs mainly via the regulation of topoisomerase gene transcription: relaxation triggers the up-regulation of gyrase and the down-regulation of topoisomerases I and IV, while hypernegative supercoiling down-regulates the expression of topoisomerase I. Relaxation affects 13% of the genome, with the majority of the genes affected located in 15 domains. Hypernegative supercoiling affects 10% of the genome, with one quarter of the genes affected located in 12 domains. However, all the above domains overlap, suggesting that the chromosome is organized into topological domains with fixed locations. Based on its response to relaxation, the pneumococcal chromosome can be said to be organized into five types of domain: up-regulated, down-regulated, position-conserved non-regulated, position-variable non-regulated, and AT-rich. The AT content is higher in the up-regulated than in the down-regulated domains. Genes within the different domains share structural and functional characteristics. It would seem that a topology-driven selection pressure has defined the chromosomal location of the metabolism, virulence and competence genes, which suggests the existence of topological rules that aim to improve bacterial fitness.

**Keywords:** DNA supercoiling, DNA topoisomerases, fluoroquinolones, global transcription, interactome, novobiocin, seconeolitsine, topological domains

## INTRODUCTION

The compaction of DNA by up to 1000-fold (Holmes and Cozzarelli, 2000) in the bacterial chromosome, or nucleoid, achieves the optimal condition under which its essential functions – replication, segregation and gene expression (reviewed by Dorman, 2013) – can be reconciled. This compaction is mediated by both the natural supercoiling of the DNA, and by the binding of nucleoid-associated proteins (NAPs) (Wang et al., 2013). NAPs form a functional network that maintains DNA topology by bending, wrapping, bridging and constraining supercoils. Although several NAPs have been characterized in the Gram-negative bacterium *Escherichia coli*, very few have been detected in Gram-positive bacteria, including the human pathogen *Streptococcus pneumoniae* (Dillon and Dorman, 2010). In bacteria, gene transcription is regulated by DNA-supercoiling. This functions as a general *cis* regulator of transcription, and can be superimposed upon other types of more specific *trans* regulatory mechanisms. *cis* regulation can also occur via promoter DNA sequences. Factors acting in *trans* include structural and regulatory proteins. NAPs (structural proteins) target a number of genes (Dillon and Dorman, 2010), while specific regulatory proteins facilitate or inhibit the interaction of RNA polymerase with specific promoter regions (Browning and Busby, 2004). The precision balance of DNA supercoiling is thus modulated by a network of self-regulating factors.

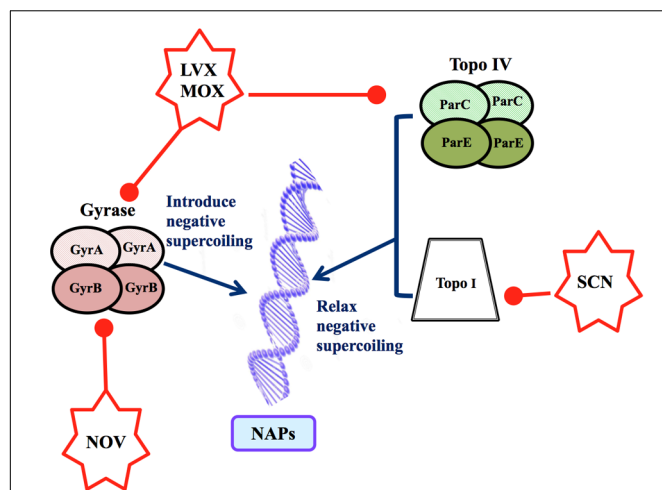
DNA topoisomerases, which are present in all bacteria, are responsible for the maintenance of DNA-supercoiling. These enzymes are classified into two types based on their DNA cleavage pattern: type I, which cleaves only one DNA strand, and type II, which cleaves both. The type II topoisomerases, gyrase and topoisomerase IV (Topo IV), are tetrameric proteins with two subunits: GyrA<sub>2</sub>GyrB<sub>2</sub> in gyrase, and ParC<sub>2</sub>ParE<sub>2</sub> in Topo IV. Supercoiling homeostasis is achieved by the competing activities of gyrase and topoisomerase I (Topo I, a type I isomerase) plus IV (Champoux, 2001); gyrase introduces negative supercoils into DNA (Gellert et al., 1976), Topo I relaxes DNA, and Topo IV both relaxes DNA and participates in chromosome partitioning (Kato et al., 1990). *S. pneumoniae* (the pneumococcus) has a relatively small genome (~2 Mb compared to ~4.6 Mb for *E. coli*) rich in AT (60%), that carries genes for all three of the above enzymes. These characteristics are shared by other pathogens of the genus *Streptococcus*, including *S. pyogenes* and *S. suis*.

*Streptococcus pneumoniae* is the primary cause of community-acquired pneumonia, meningitis, bacteremia, and otitis media in children. Worldwide, 1 million children under 5 years of age die every year of pneumococcal infections (World Health Organization, 2007). The use of the pneumococcal 7-valent conjugate vaccine, which covers the serotypes most often associated with resistance to antibiotics, has achieved a decline in the incidence of invasive pneumococcal disease (Whitney et al., 2003; Kyaw et al., 2006) and a reduction in penicillin resistance rates (Kyaw et al., 2006; Pilishvili et al., 2010). However, serotypes not included in

the vaccine soon emerged, highlighting the limitations of anti-pneumococcal prophylaxis (Moore et al., 2008; Fenoll et al., 2009).

The post-genomic age is beginning to provide answers to questions regarding how chromosomes are topologically organized, and how this organization influences bacterial evolution. Several degrees of organization in bacterial chromosomes have been observed, based on size (for a recent review see Badrinarayanan et al., 2015). Macrodomains are found at the megabase-size range. *E. coli*, for example, has four macrodomains: Ori (origin of replication), Ter (terminus of replication), Left, and Right, plus two less-structured regions flanking the Ori macrodomain (Espeli et al., 2008). Macrodomains may be maintained by specific proteins, such as the macrodomain Ter proteins (MatPs) that bind, as the name suggests, to specific sites in the Ter macrodomain (Dupaigne et al., 2012). However, no such proteins stabilizing the other macrodomains have been identified, and MatP proteins are found only in enteric bacteria. Non-homologous proteins may therefore take on similar roles in other bacteria. Supercoiling domains are found at the kilobase range. These are isolated loops that coil up around themselves; proteins at their bases help to topologically isolate the looped DNA. These loops were initially detected in electron micrographs of lysed *E. coli* cells (Kavenoff and Bowen, 1976). Later studies estimated the number of supercoil domains by assessing the numbers of nicks required to fully relax the chromosome. From these experiments it was estimated that the *E. coli* chromosome contains about 40 domains of around 100 kb (Worcel and Burgi, 1972; Sinden and Pettijohn, 1981). Studies in *Caulobacter crescentus* suggested domains ranging in length from 30 to 420 kb (Le et al., 2013). In *Salmonella enterica*, these domains were estimated to be 20 kb long by taking into account the site-specific recombination events that occurred between chromosomal sites distant from one another (Higgins et al., 1996). Later, transcriptional data predicted sizes of ~10 kb for *E. coli* (Postow et al., 2004). Controversy regarding the size and definition of domains remains, perhaps as a consequence of the different methods being used in their calculation.

The availability of drugs against all the topoisomerases of *S. pneumoniae* (Figure 1) has helped in determining the existence of chromosomal domains. This review summarizes the transcriptomic alterations induced by these agents, and how these changes can be interpreted to provide definitions of the chromosome domains in this bacterium. Changes induced by the clinically used fluoroquinolones (FQs) levofloxacin (LVX), and moxifloxacin (MOX) are first considered, followed by those that occur concomitantly with a global change in supercoiling, as induced by novobiocin (NOV, an inhibitor of the gyrase B subunit) and seconeolitsine (SCN, an inhibitor of Topo I). Overall, these studies reveal the *S. pneumoniae* genome to be organized into topology-reacting gene clusters, or supercoiling domains. The conservation of the location of these domains in the *Streptococcus* genus, and their enrichment for specific functions, suggests the existence of topological rules that aim to improve fitness via tight physiological feedback.



**FIGURE 1 |** Factors determining the topology of the *Streptococcus pneumoniae* chromosome. The level of supercoiling is controlled by three DNA topoisomerases. Gyrase (GyrA<sub>2</sub>GyrB<sub>2</sub>) is inhibited by LVX, MOX (which inhibit GyrA) and NOV (which inhibits GyrB). Topo IV (ParC<sub>2</sub>ParE<sub>2</sub>) is inhibited by LVX and MOX (which inhibit ParC). Topo I, a monomer, is inhibited by SCN. The topological organization of the chromosome depends on the level of DNA supercoiling and on the presence of NAPs.

## CONTROL OF TRANSCRIPTION BY LOCAL CHANGES IN SUPERCOILING

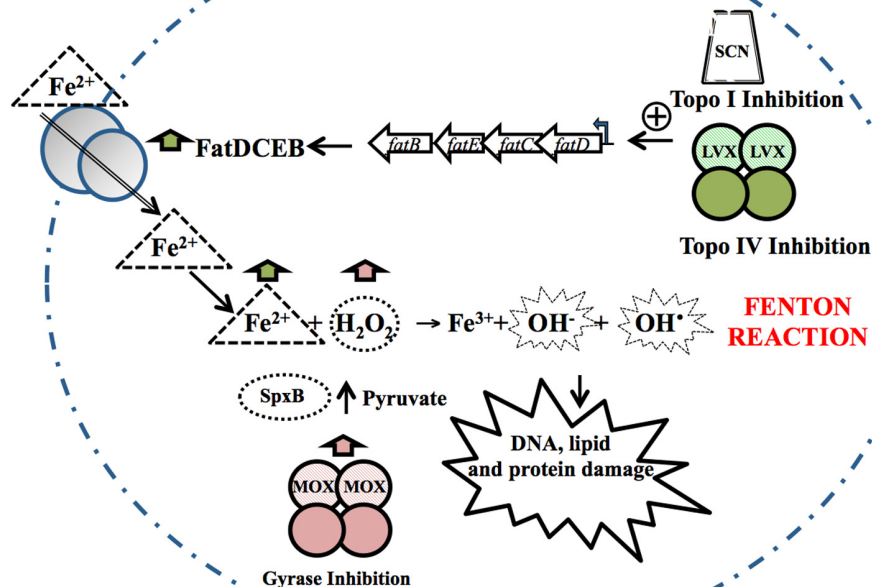
Strains of *S. pneumoniae* resistant to antibiotics that act on the cell wall (beta-lactams) and on protein synthesis (macrolides) have proliferated in the last 30 years (Jacobs et al., 2003; Liñares et al., 2010). Consequently, pneumococcal infections are nowadays fought with LVX and MOX, which inhibit DNA topoisomerases. FQs target the type II DNA topoisomerases gyrase and Topo IV. Their mechanism of action involves the formation of DNA-FQ-topoisomerase complexes, which sterically inhibit replication and transcription and the subsequent generation of detrimental double-stranded DNA breaks (Drlica et al., 2008). Bacterial survival depends on the resolution of these breaks. Reactive oxygen species (ROS), such as superoxide anions, hydrogen peroxide and hydroxyl radicals contribute to FQ-mediated cell death via a protein synthesis-dependent pathway (Wang et al., 2010). This observation is consistent with the general model explaining the lethality of bactericidal antibiotics, which attributes a role to ROS generated via the Fenton reaction. The original reports supporting this model based their conclusions on the use of microarrays to study the transcriptional response to the inhibition of *E. coli* GyrA by an FQ or the peptide toxin CcdB. Under these conditions, global transcription was altered. In addition to the up-regulation of SOS damage response genes, genes related to superoxide stress, iron-sulfur cluster synthesis and iron uptake were up-regulated too (Dwyer et al., 2007). ROS production was also observed with a variety of bactericidal antibiotic families, in addition to FQs, each with a different intracellular target (reviewed by Dwyer et al., 2015). However, the intervening pathways lying between

the initial antibiotic-target interaction and ROS formation have yet to be fully characterized.

The treatment of *S. pneumoniae* with FQs involves causing double-stranded breaks in the bacterial chromosome (Ferrándiz et al., 2016b), and as in other bacteria this requires active protein synthesis (Brito et al., 2017). Treatment with LVX or MOX (Ferrándiz and de la Campa, 2014; Ferrándiz et al., 2016b) is reported not to alter the level of global supercoiling. Nor are changes in supercoiling observed in *E. coli* exposed to oxolinic acid (Snyder and Drlica, 1979), although changes have been observed in the latter after treatment with the FQ norfloxacin (Peter et al., 2004). These differences might be attributable to species-dependent affinities of each drug for Topo IV or gyrase. For instance, Topo IV is the primary target of most FQs in Gram-positive bacteria, including *S. pneumoniae*, with gyrase a secondary target (Janoir et al., 1996; Muñoz and de la Campa, 1996; Tankovic et al., 1996; Fernández-Moreira et al., 2000). In contrast, in Gram-negative bacteria, including *E. coli*, gyrase is the primary target. At the LVX concentrations used in *S. pneumoniae* experiments, only Topo IV would have been inhibited, and no global change in supercoiling would be expected. However, at the MOX concentrations used, both gyrase and Topo IV would have been inhibited, suggesting that the inhibition of their opposing activities preserved the net level of supercoiling. Nevertheless, local topological changes are predictable in both cases and these would produce alterations in the transcriptome. Indeed, FQs induce a transcriptional response in *S. pneumoniae*, in which the differentially expressed genes (DEGs) account for 5.2 and 6.5% of the genome for LVX and MOX, respectively. In this bacterium, which lacks a proper SOS-like system, activation of the competence regulon has been reported with both FQs (Ferrándiz and de la Campa, 2014; Ferrándiz et al., 2016b), supporting the idea that competence is a general stress response in *S. pneumoniae* (Prudhomme et al., 2006). In addition, both LVX and MOX induce transcriptional alterations, which, although different, ultimately stimulate the Fenton reaction, increasing ROS accumulation and contributing to cell death (Ferrándiz and de la Campa, 2014; Ferrándiz et al., 2016b). Although *S. pneumoniae* is a facultative anaerobe, the increased lethality of FQs mediated by an increase in ROS fits with the antibiotic lethality model proposed for aerobic bacteria (Dwyer et al., 2007, 2014; Kohanski et al., 2007; Wang and Zhao, 2009). Via local supercoiling changes, the response to LVX specifically triggers the up-regulation of the *fatDCEB* operon. This causes an increase in intracellular iron, and in turn, a shift in the Fenton reaction toward the production of hydroxyl radicals. With MOX, the response leads to the up-regulation of the glycolytic pathway, with a noticeable increase in pyruvate and a subsequent increase in hydrogen peroxide (Figure 2). The different alterations in the patterns of gene expression induced by LVX and MOX are due to local changes in supercoiling, which are dependent on whether Topo IV (LVX) or both Topo IV and gyrase (MOX) are inhibited.

Since both Topo IV and gyrase produce double-stranded breaks in the DNA when bound to FQs, the differential transcriptional alterations caused by these drugs might also





**FIGURE 2 |** Oxidative damage cell death pathway. The inhibition of Topo IV by LVX, or of Topo I by SCN, causes a local increase in supercoiling, resulting in the up-regulation of the *fatDCEB* operon. The consequent increase in the iron transporter it encodes causes an increase in intracellular  $Fe^{2+}$ . MOX alters the transcriptome, up-regulating genes from several metabolic pathways involved in the production of pyruvate. Pyruvate can then be converted by pyruvate oxidase (SpxB) into hydrogen peroxide ( $H_2O_2$ ).  $Fe^{2+}$  and  $H_2O_2$  are the substrates of the Fenton reaction. The Fenton reaction renders hydroxyl radicals, which oxidatively damage DNA, proteins and lipids. Taken from Ferrándiz and de la Campa (2014), with modifications.

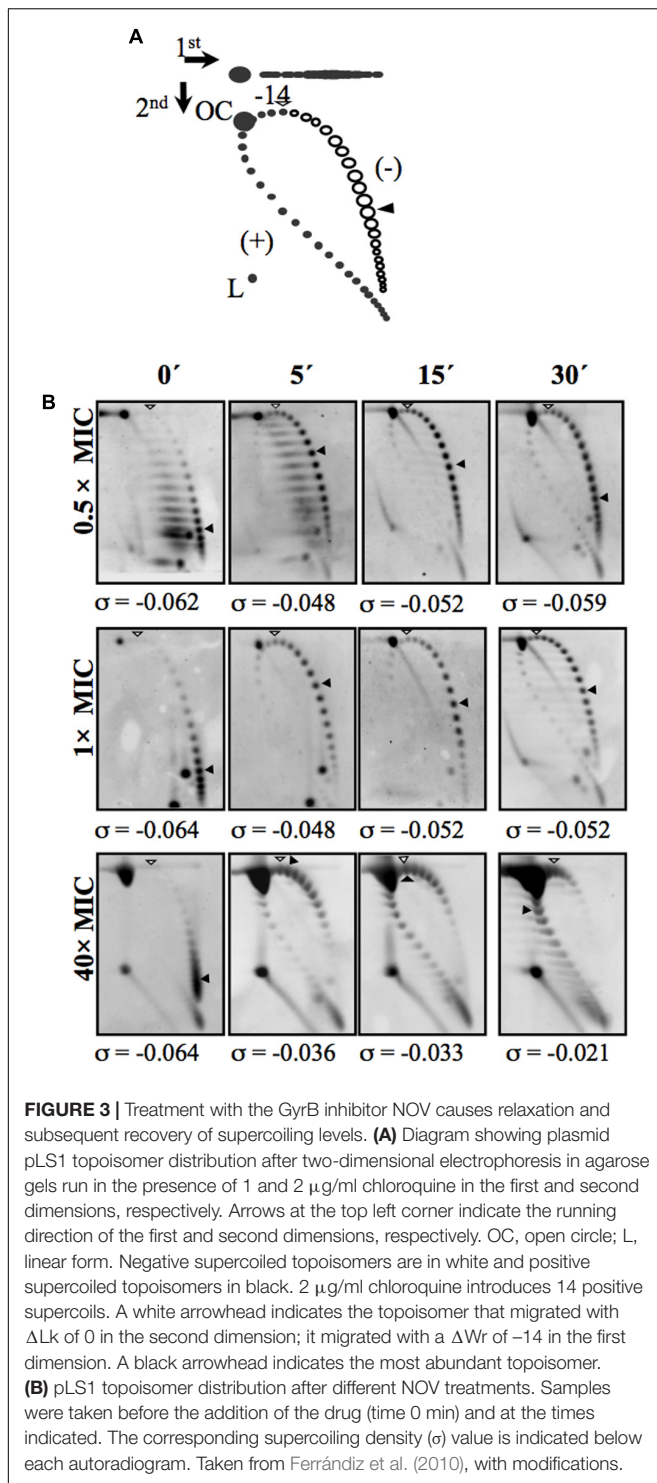
be related to subtle, yet important, differences in sequence recognition (Leo et al., 2005), which are themselves affected by DNA supercoiling and bending (Arnoldi et al., 2013). Sequence recognition mediated by local supercoiling levels might explain the unique distribution of genes affected by LVX or MOX. In addition, the location of FQ-topoisomerase complexes relative to the replication forks, which is different for gyrase and Topo IV (Postow et al., 2001), may be involved in their different transcriptional outcomes.

## CONTROL OF TRANSCRIPTION BY GLOBAL CHANGES IN SUPERCOILING

### Response to Relaxation Caused by the Inhibition of Gyrase

The homeostatic control of supercoiling was first described in *E. coli*. In this bacterium, the transcription of *topA* (which codes for Topo I) was found to decrease under DNA relaxation (Tse-Dinh, 1985), while that of *gyrA*, and *gyrB* (which code for the two gyrase subunits) were found to increase (Menzel and Gellert, 1983, 1987a,b). An increase in gyrase expression in response to relaxation has also been observed in *Streptomyces* and *Mycobacterium* (Thiara and Cundliffe, 1989; Unniraman

et al., 2002). However, in *Staphylococcus aureus*, treatment with NOV affects the transcription of the gyrase genes but not of *topA* (Schroder et al., 2014). In *S. pneumoniae*, treatment with NOV was also found to increase the transcription of gyrase genes, and diminish the expression of Topo I and Topo IV. In addition, global relaxation followed by a recovery of the native level of supercoiling was observed at low drug concentrations (Ferrándiz et al., 2010). The distribution of topoisomers in plasmid pLS1 (Stassi et al., 1981) was used to estimate the chromosomal superhelical density ( $\sigma$ ), and returned a mean value of about  $-0.06$  (Figure 3), which is within the range reported for the *E. coli* chromosome (Deng et al., 2005). At subinhibitory NOV concentrations ( $0.5 \times$  MIC), a transcriptomic response allowed the restoration of the native level of supercoiling after an initial relaxation causing a  $\sigma$  variation of 23%. A similar effect was observed at  $1 \times$  MIC. However, higher concentrations of NOV increased the degree of relaxation with no further restoration of supercoiling, compatible with the saturation of the homeostatic capacity that results in the inhibition of cell division. The range of  $\sigma$  variation permitting homeostatic recovery of the supercoiling observed in *S. pneumoniae* is in agreement with the estimated  $\pm 20\%$  variation compatible with normal cell growth in *E. coli* (Drlica, 1992). Supercoiling recovery in the pneumococcus occurred after the up-regulation of the gyrase genes *gyrA* and *gyrB* and the down-regulation of the Topo I



(*topA*) and Topo IV (*parEC*) genes (Ferrándiz et al., 2010). In *E. coli*, the expression of the gyrase and Topo I genes is also mediated by NAPs, which affect DNA supercoiling (Travers and Muskhelishvili, 2005; Vora et al., 2009). However, these regulatory mechanisms may not function in *S. pneumoniae* for which NAP scarcity is predicted, and which certainly lacks most

of the NAPs found in *E. coli*. Thus, supercoiling maintenance in *S. pneumoniae* appears to depend mainly on the regulation of topoisomerase transcription.

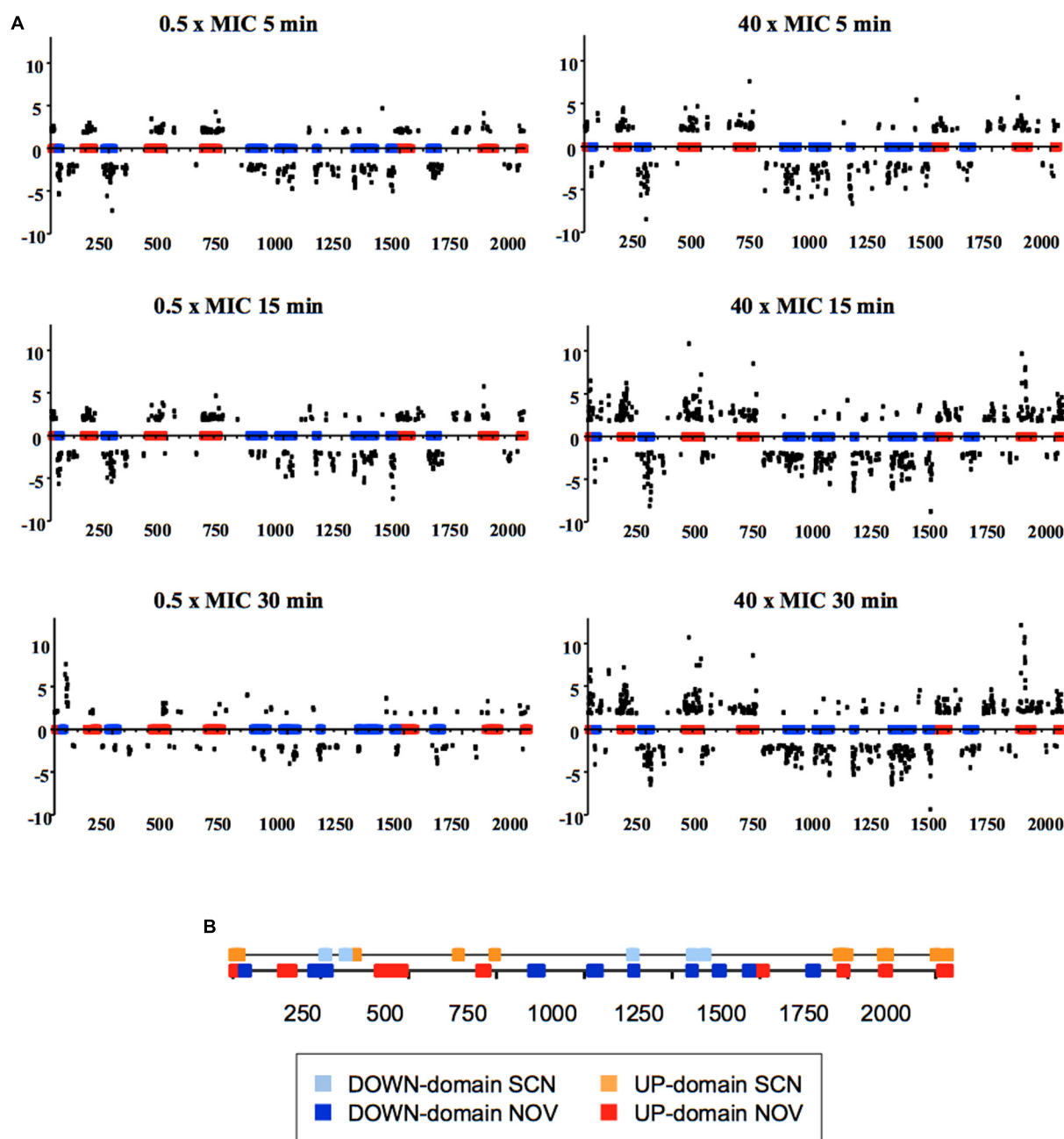
### The Transcriptional Response to DNA Relaxation Involves Topology-Reactive Gene Clusters

The modulation of the expression of topoisomerase genes in *S. pneumoniae* is part of a global genome response (Ferrándiz et al., 2010). At subinhibitory concentrations, i.e., under physiological conditions, and short treatment times (5 and 15 min), DEGs were found to account for about 13% of the genome. An attenuation in the response at 30 min was observed, the number of DEGs being reduced to account for just 5.7% of the genome (Figure 4A), reflecting the recovery of supercoiling (Figure 3). Some 13% of the pneumococcal genome was therefore involved in the cellular response to moderate relaxation, allowing the recovery of the initial level of supercoiling. At fully inhibitory concentrations, the proportion of the genome covered by DEGs increased with time, from 14.4% at 5 min to 24% over longer periods (Figure 4). This agrees with the inhibition of cell division and with the continuous relaxation of the DNA (Figure 3). This proportion of the genome covered by DEGs upon relaxation is larger than in other bacteria. In Gram-negative bacteria, DEGs were found to account for 7% of the genome in *E. coli* [as determined using both gyrase inhibitors and gyrase thermosensitive mutants (Peter et al., 2004)], and for 8% in *Haemophilus influenzae* [as determined using NOV (Gmüender et al., 2001)]. In Gram-positive *Staphylococcus aureus*, treatment with NOV affected the transcription of 11% of the genome (Schroder et al., 2014).

It should be noted that the transcriptomic response to relaxation in *S. pneumoniae* involves topology-reactive gene clusters, or domains, that show coordinated up- or down-regulation. A total of 15 clusters have been detected, corresponding to 37% of the genome (Figure 4) (Ferrándiz et al., 2010). The sizes of these clusters varies from 14.6 to 85.6 kb (mean  $\pm$  SD:  $51.8 \pm 21.8$ ) and they contain 15–43 responsive genes (mean  $\pm$  SD:  $28 \pm 9$ ). They also include more than 68% of the DEGs. This has allowed topological clusters to be identified in which gene co-regulation is clearly more complex than would be expected simply from the number of genes in operons. In addition, the direction of transcription of the DEGs showed no preference for leading or lagging strands, providing additional evidence that topological control is structurally dependent.

The AT content over the genome correlates with domain location, and is higher in up-regulated (UP) than in down-regulated (DOWN) domains. These results suggest that the relaxation of DNA in AT-rich (ATr) regions favors the access of RNA polymerase to their promoters. On the contrary, a low AT content in DOWN clusters obstructs the access of RNA polymerase. Enrichment in the AT content of the region from positions -800 to +200 of genes up-regulated under relaxation has been reported in *E. coli* (Peter et al., 2004).

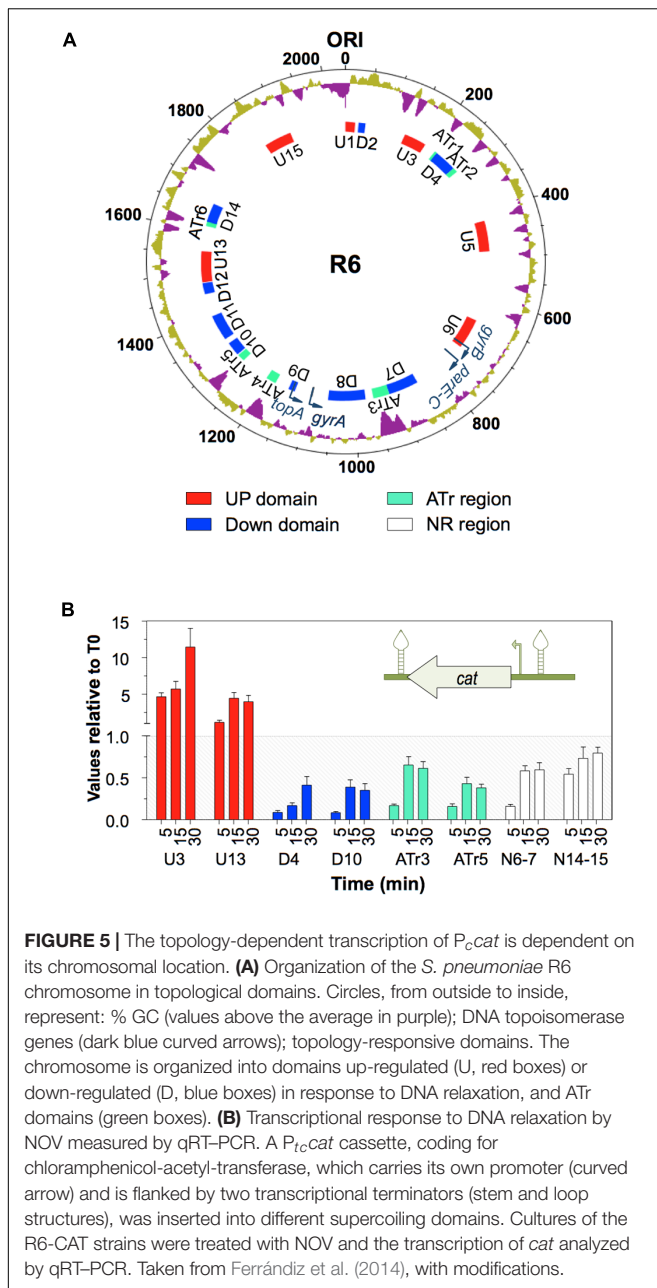
The organization of the *S. pneumoniae* chromosome into domains was further confirmed by the introduction of a *cat* heterologous gene cassette into the different types of domain (Figure 5A) (Ferrándiz et al., 2014). In response to



**FIGURE 4 |** Global transcriptomic responses of *S. pneumoniae* to relaxation with NOV. **(A)** The relative fold variation of each gene is represented against the 3' location of each open reading frame in the *S. pneumoniae* R6 chromosome (bases 1 to 2,038,615). Boxes indicate the transcriptional clusters: up-regulated in red, down-regulated in blue. **(B)** Localization of topological clusters detected as a result of treatment with either SCN or NOV. Taken from Ferrández et al. (2010, 2016a), with modifications.

relaxation with NOV, the transcription of *cat* was dependent on its chromosomal location, being up-regulated when located in UP domains, down-regulated when located in DOWN domains, and showing almost no changes when located in the non-regulated (NR) domains (Figure 5B). This all supports the idea that the chromosome is organized into topological domains that are reactive to interference in the

supercoiling status. These results contrast, however, with those obtained in *E. coli*, in which the 306 DEGs were not only functionally diverse but widely dispersed throughout the chromosome (Peter et al., 2004), and with results obtained for *Staphylococcus aureus*, in which NOV-responsive genes were randomly distributed throughout the chromosome (Schroder et al., 2014).



## Response to Hypernegative Supercoiling Caused by the Inhibition of Topo I

The negative supercoiled state is the natural state of DNA homeostatic equilibrium in many bacteria. However, hypernegative supercoiling has been reported in *E. coli* *topA* mutants. With the exception of the *topA10* mutant, all have acquired compensatory mutations in the gyrase genes (DiNardo et al., 1982). The *topA10* mutant shows a notable 22% increase in negative supercoiling (Pruss et al., 1982), which probably represents the limit viable cells can afford in the long term. The inhibition of Topo I would produce greater hyper-supercoiling. Topo I plays an essential role in transcription, given its physical

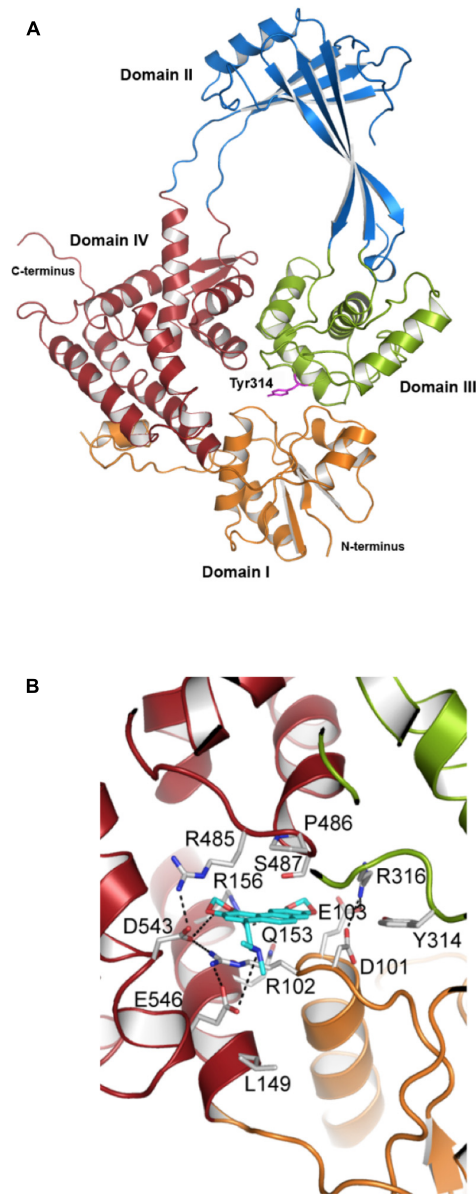
interaction with RNA polymerase (Cheng et al., 2003). During transcription, hypernegative supercoiling occurs behind the RNA polymerase, leading to RNA-DNA hybrid (R-loop) stabilization (Drolet, 2006). Topo I relaxes this supercoiling and prevents R-loop formation (Drolet et al., 1994; Phoenix et al., 1997; Masse and Drolet, 1999), allowing transcription to continue. Thus, the effects of hypernegative supercoiling in transcription depend directly on the activity of Topo I.

However, Topo I-targeting compounds are extremely scarce. Cheng et al. (2007) identified an alkaloid, which, although it inhibits the activity of *E. coli* Topo I, did not inhibit cell growth significantly. Our group discovered a new inhibitor of *S. pneumoniae* Topo I, SCN, which inhibits its relaxation activity at concentrations equivalent to those that inhibit cell growth. The modeling of pneumococcal Topo I, based on the crystal structure of the *E. coli* enzyme (Figure 6), and docking to SCN, revealed strong interactions between the drug and the DNA-binding site of Topo I to correlate with the inhibitory effect observed (García et al., 2011).

Our group was the first to use SCN in studies of the transcriptomic response to hypernegative supercoiling in bacteria (Ferrándiz et al., 2016a). The viability of *S. pneumoniae* and the increase in supercoiling is affected by SCN in a concentration-dependent manner (Figure 7). Treatment with 6  $\mu$ M SCN produced a peak  $\sigma$  increase of 41% at 5 min, which later recovered. Treatment with 8  $\mu$ M SCN resulted in higher and longer lasting increases in the  $\sigma$  value, with partial recovery after 120 min. These results show that treatment with subinhibitory SCN concentrations permit the recovery of peak  $\sigma$  increases of up to 41% without affecting cell viability. This tolerance to increases in supercoiling levels is greater than the 25% observed for DNA relaxation upon NOV treatment (Figure 8A) (Ferrándiz et al., 2010), and indicates that *S. pneumoniae*, and very likely genetically related bacteria, are naturally more tolerant to hypernegative supercoiling than to hyper-relaxation. Similarly, the results of experimental evolution assays with *E. coli* revealed increasing supercoiling (associated with mutations in *topA*) to increase bacterial fitness (Crozat et al., 2005). A similar homeostatic mechanism allowing increased negative supercoiling might also exist in bacteria with reverse gyrase. These bacteria keep DNA in a slightly overwound state to protect their genome from heat damage (Ogawa et al., 2015).

The transcription levels of *topA* in *S. pneumoniae* at subinhibitory concentrations of SCN or NOV (which allow for cell growth and the recovery of supercoiling) show a good correlation with the induced variation in  $\sigma$  (Figure 8B). The regulation of *topA* therefore plays a fundamental role in the recovery of supercoiling levels. The variations seen in *topA* expression were, however, only part of a global transcriptomic response. Treatment with subinhibitory concentrations of SCN (8  $\mu$ M, 0.5 $\times$  MIC) generated a two-stage transcriptomic response: (i) early response and (ii) recovery. The former, which represents an active response against sharply increased supercoiling, was observed at 5 and 15 min of treatment, and involved about 11% of the genome. During recovery, only about 2% of the genome was involved at 30 min. In the early response, transcriptional variations also occurred in clusters, with DEGs





**FIGURE 6 |** Structural modeling of the interaction of *N*-methyl SCN with *S. pneumoniae* topoisomerase I. **(A)** Modeling of the 67 kDa fragment of Topo I, showing domains I–IV and the catalytic Tyr314. **(B)** *N*-methyl-SCN (in blue) bound to the nucleotide-binding site of Topo I. Hydrogen bonds and salt-bridge interactions are indicated by dashed lines. Taken from García et al. (2011), with modifications.

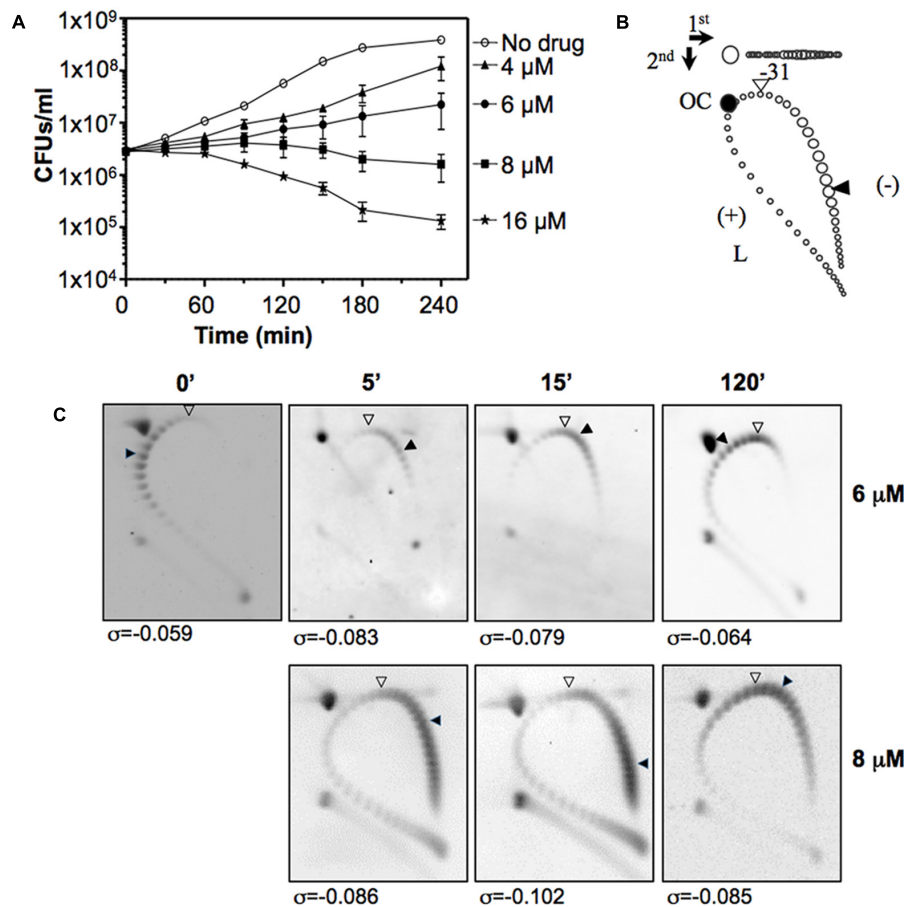
grouping into topologically sensitive domains. The average size of a SCN cluster is  $14.0 \pm 7.6$ , similar to the 10 kb *E. coli* domains predicted using transcriptional data (Postow et al., 2004). Although the NOV and SCN clusters are not identical, their position in the chromosome nearly overlap (**Figure 4B**) – an unexpected finding given the opposing nature of DNA relaxation and supercoiling. These results support the idea that the chromosome is divided into topological domains with fixed locations.

## Regulation of DNA Topoisomerase Gene Transcription

In *E. coli*, several NAPs are involved in the regulation of topoisomerases. One such NAP is the FIS protein, which regulates the expression of genes coding for the subunits of gyrase (Schneider et al., 1999), Topo I (Weinstein-Fischer and Altuvia, 2007), and the genes coding for other NAPs involved in DNA supercoiling (Claret and Rouviere-Yaniv, 1996; Falconi et al., 1996; Grainger et al., 2008). In addition, two further NAPs, FIS, and H-NS proteins control both the level of supercoiling and global transcription (Blot et al., 2006; Marr et al., 2008). The corresponding situation in *S. pneumoniae*, which lacks these NAPs, seems to be much simpler.

The transcription of *gyrB* and *topA* in *S. pneumoniae* is regulated by their strategic chromosomal location in topological domains, since the expression driven by their promoters differs whether they are located in their natural chromosomal locations or in a replicating plasmid (Ferrándiz et al., 2014). Transcriptional fusions of these promoters to a reporter gene in plasmid pLS1 have been measured after DNA relaxation induced by NOV. As expected, relaxation caused down-regulation of *topA* and up-regulation of *gyrB* when the genes were located in their native chromosomal sites (DOWN9 for *topA* and UP6 for *gyrB* in **Figure 5A**). However, transcription from both promoters in the plasmid fusions was down-regulated. These results indicate that both *topA* and *gyrB* are under supercoil-mediated regulation, and that the plasmid behaves as a DOWN domain. This may serve to neutralize the high copy number of the plasmid genes and/or favor their replication.

In contrast, the Topo IV genes (*parE* and *parC*) and *gyrA* are located in NR domains, and their expression depends on specific regulatory signals located in the promoter region. The expression of the Topo IV genes from their common promoter (Balsalobre and de la Campa, 2008) is equivalent in their natural chromosomal location and in plasmids (Ferrándiz et al., 2014). With respect to the *gyrA* gene, its upstream region ( $P_{gyrA126}$ , nt –126 to +1 in **Figure 9A**) shows an intrinsic DNA curvature (Balas et al., 1998). This was fused to *cat* and cloned into plasmid pLS1, and the curvature either eliminated by a 5 bp insertion ( $P_{gyrA126Pae}$ ) or by a 5 bp deletion ( $P_{gyrA121Pae}$ ), and a direct correlation observed between *cat* expression and the curvature under basal conditions (the specific activity of the  $P_{gyrA126}$  fusion was ~3-fold higher than that recorded for plasmids lacking the curvature). This shows that the curvature behaves as an activator *per se*, providing better recruitment of either the RNA polymerase complex or specific regulatory proteins. The role of curvatures as regulators of transcription has previously been established in bacteria (Pérez-Martín et al., 1994), including *S. pneumoniae* (Pérez-Martín and Espinosa, 1991). In addition, the transcription levels from the chromosomal  $P_{gyrA}$  and the  $P_{gyrAcat}$  fusions in plasmids in the presence of NOV have been determined. While in the plasmid carrying the wild-type promoter ( $P_{gyrA126}$ ) the up-regulation of *cat* was similar to that of the chromosomal *gyrA*, down-regulation of *cat* was observed in the plasmids lacking the curvature (**Figure 9B**). These results suggest that the signals regulating *gyrA* transcription are included within the above-mentioned 126 nt region, and that bending is a key element



**FIGURE 7 |** Seconeolitsine affects cell viability and induces hypersupercoiling. **(A)** Viability of R6 (pLS1) in a medium containing different SCN concentrations. Samples were taken before the addition of the drug (zero time), and at the indicated times were plated on drug-free agar medium. **(B)** Diagram showing topoisomer distribution in plasmid pLS1 subjected to 2D-agarose gel electrophoresis run in the presence of 5 and 15  $\mu$ g/ml chloroquine in the first and second dimensions, respectively. OC, open circle; L, linear form. A white arrowhead points to the topoisomer migrating with a  $\Delta Lk$  of 0 in the second dimension; it had a  $\Delta W_r$  of -31 in the first dimension. A black arrowhead points to the more abundant topoisomer. **(C)** Distribution of pLS1 topoisomers in 2D-gels after treatment with the indicated SCN concentrations. The supercoiling density ( $\sigma$ ) values are indicated. Taken from Ferrándiz et al. (2016a), with modifications.

for its regulation under relaxation by acting as a sensor of the supercoiling level.

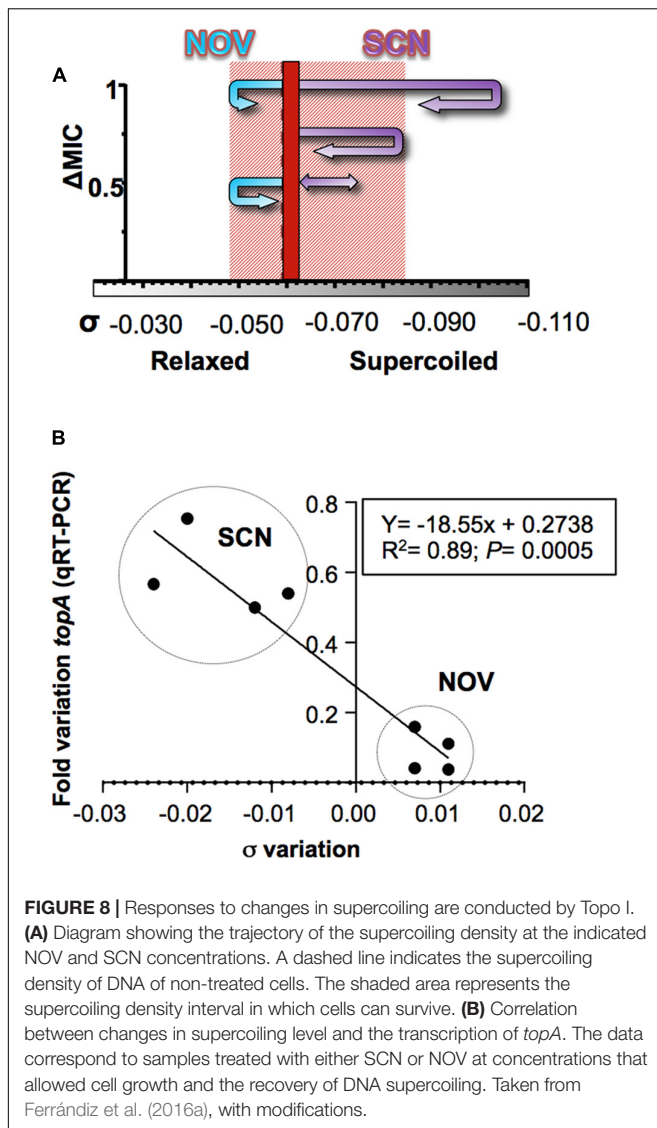
Chromatin immunoprecipitation experiments using antibodies directed against the pneumococcal GyrA subunit and Topo I (Ferrándiz et al., 2016a) have shown  $P_{gyrA}$  to recruit Topo I, but not gyrase (Figure 9C). The region to which Topo I binds includes the -35 and extended -10 boxes on  $P_{gyrA}$ , plus the DNA curvature (Balas et al., 1998). Thus, Topo I, the transcription of which is regulated by supercoiling levels, appears to be the key factor regulating *gyrA* expression.

## EVOLUTIONARY PRESSURE DRIVES THE ORGANIZATION OF THE CHROMOSOME INTO DOMAINS

### Domain Conservation in Streptococci

Gene order in bacterial chromosomes surpasses the level of the operon (Lathe et al., 2000; Reams and Neidle, 2004). As explained

above, and based on its transcriptome under DNA relaxation, the chromosome of *S. pneumoniae* R6 appears to be organized into four types of topological domains: UP, DOWN, NR, and ATr. The analysis of 12 *S. pneumoniae* complete genome sequences has revealed the conservation of the UP and DOWN domains (Figure 10). The gene-lack index (number of genomes in which a gene is absent divided by the total number of genomes) revealed lower values for the UP (1.51) and DOWN (1.65) domains than the genome average (1.91). However, ATr domains have high gene-lack indices (average 4.66), suggesting extensive gene interchange in these domains. To study the conservation of domains, normalized location dispersion indices (nLDI: values that quantify the position deviation of a given gene with respect to the Ori, and relative to homologs in several genomes (Martín-Galiano et al., 2017)) were calculated across *S. pneumoniae* genomes; the values returned were very small since synteny is highly conserved in this species. The same was then calculated for representative strains of 25 species of *Streptococcus* in order to detect distinguishing differences. The conservation of



**FIGURE 8 |** Responses to changes in supercoiling are conducted by Topo I. **(A)** Diagram showing the trajectory of the supercoiling density at the indicated NOV and SCN concentrations. A dashed line indicates the supercoiling density of DNA of non-treated cells. The shaded area represents the supercoiling density interval in which cells can survive. **(B)** Correlation between changes in supercoiling level and the transcription of *topA*. The data correspond to samples treated with either SCN or NOV at concentrations that allowed cell growth and the recovery of DNA supercoiling. Taken from Ferrándiz et al. (2016a), with modifications.

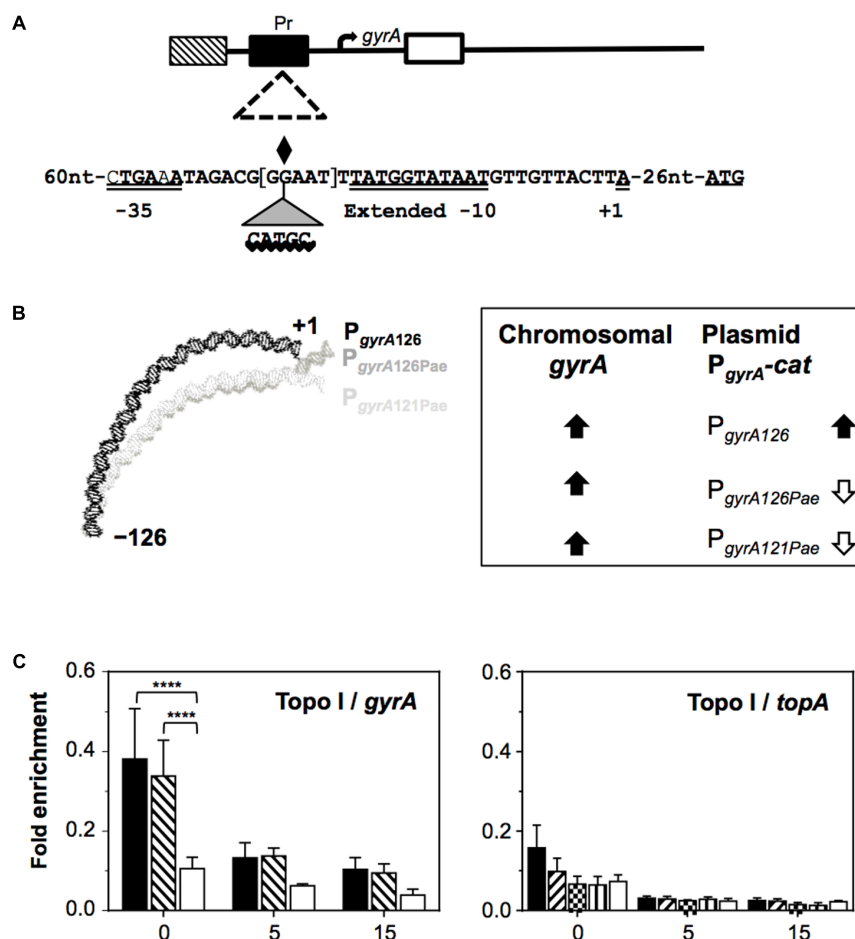
*S. pneumoniae* domains across these *Streptococcus* representatives was then determined. Two assumptions were made: (i) that the gene order is relatively conserved, as seen in gamma-proteobacteria (Sobetzko et al., 2012), and (ii) that chromosomal topology is conserved, given that species share core gene pools (Lefebvre and Stanhope, 2007), similar genome lengths, and a similar AT content. Similar approaches have been followed to examine chromosomal patterning in other bacteria (Wright et al., 2007; Khedkar and Seshasayee, 2016). In *S. pneumoniae*, 571 genes (28.0%) had nLDI values of  $<1$ , which indicates they tend to locate to positions more stable than the average for maintained homologs (Martín-Galiano et al., 2017). Several genes from the UP and DOWN domains were present in most streptococci at equivalent positions. The greatest position conservation was observed in 40 genes near the Ori, indicating strong topological pressure to maintain functionalities in this region. Genes near the Ori have high copy numbers (Slager and Veening, 2016) and show a peculiar pattern of NAP binding (Sobetzko et al., 2012).

Moreover, seven clusters with conserved positions were detected for NR genes, and named pcNR domains (position-conserved Non-Regulated domains). Most of the remaining NR genes were organized into 14 domains ( $\geq 10$  genes) termed pvNR domains (position-variable Non-Regulated). ATr regions accounted for 13 domains (Figure 11). Strikingly, the pcNR domains appeared symmetrically located at regular intervals ( $\sim 200, 400$ , and  $800$  kb) on both sides of the Ori and were interleaved between UP, DOWN, and pvNR domains (Figure 11A). The size of these domains appeared compatible with the  $100$  kb lengths estimated for them using different techniques (Worcel and Burgi, 1972; Sinden and Pettijohn, 1981; Le et al., 2013). This suggests a potential higher-order macrostructural unit above the domain level controlling the genetic stability and plasticity required to face new environments (Rocha, 2004a).

## Levels of Protein Expression and Essentiality of the Domains

The transcriptomes of exponentially growing cultures (Ferrándiz et al., 2016a,b) showed the pcNR domain transcription level to be higher than that of the ATr domains (Figure 11B). Two factors contribute to these transcriptional differences. First, long repeat sequences (BOX, RUP, and SPRITE) (Croucher et al., 2011), which are associated with the repression of transcription, are few in pcNR domains, and second, the codon adaptation index (CAI), which is related to the translation rate and mRNA levels (Martín-Galiano et al., 2004), is high in pcNR domains (Martín-Galiano et al., 2017). Gene location also affects protein levels (Ochman et al., 2000; Rocha, 2004b), a pattern associated with the distance to the Ori. Genes at the Ori are doubly represented with respect to genes at the Ter in *E. coli* during exponential growth (Chandler and Pritchard, 1975). Accordingly, the relocation of genes coding for ribosomal proteins and the RNA polymerase alpha subunit to positions distant to the Ori, reduces their transcription rates, which was associated with slower growth in *Vibrio cholerae* (Soler-Bistue et al., 2015). Similarly, in *Salmonella typhimurium*, genes relocated near the Ori are expressed more strongly than those relocated near the Ter (Schmid and Roth, 1987). The regular positioning of strongly expressed genes may mark the limits of domains, as reported for *Caulobacter crescentus* (Le et al., 2013).

The fraction of essential genes, as determined by Tn-seq (van Opijnen and Camilli, 2012), is notably higher in pcNR domains than in the other domains (Figure 12A). The co-localization of essential genes beyond randomness has also been reported for *Bacillus subtilis* and *E. coli* (Fang et al., 2005), perhaps because clustering makes genomes more resistant to deletions (Fang et al., 2008). The number of pcNR genes in the lagging strand was 15.6%, significantly lower than the average in the remaining *S. pneumoniae* genome (22.3%). This would reduce the chances of collision between DNA and RNA polymerases, resulting in the discontinuation of transcription (French, 1992). Essential genes also tend to be more strongly expressed (Rocha and Danchin, 2003), as confirmed for pneumococcal pcNR genes. Essential gene clustering at regular intervals, and not affected by topological stress as defined for pcNR, appears



**FIGURE 9 |** Control of the transcription of DNA topoisomerase genes by supercoiling. **(A)** Representation of the *gyrA* coding region and of the regions tested in chromatin immunoprecipitation, showing the sequences of the wild-type  $P_{gyrA126}$ ,  $P_{gyrA126Pae}$ , and  $P_{gyrA121Pae}$  derivatives. The -35 and extended -10 boxes, the nucleotide at which transcription is initiated (+1), the center of the intrinsic DNA curvature (diamond), and the location of the inserted CATGC sequence that creates a *PaeI* restriction site, are all indicated. The five nucleotides deleted in  $P_{gyrA121Pae}$  are in brackets. **(B)** The relaxation-induced up-regulation of  $P_{gyrA}$  depends on intrinsic bending: curvature prediction and results obtained from qRT-PCR analysis. **(C)** Recruitment of Topo I to *topA* and *gyrA* upstream sequences. Exponentially growing cells were subjected to chromatin immunoprecipitation using anti-Topo I antibodies; the pulled-down DNA was subsequently analyzed by qPCR. The graphs show the pulldown efficiency (ChIP-DNA/input DNA) for each primer pair. Values are the average  $\pm$  SD of three independent replicates. \*\*\*\* $P < 0.0001$ . Taken from Ferrández et al. (2014, 2016a), with modifications.

to reflect a favorable “supercoiling environment” for protein expression.

## The Different Domains Contain Genes with Different Functions

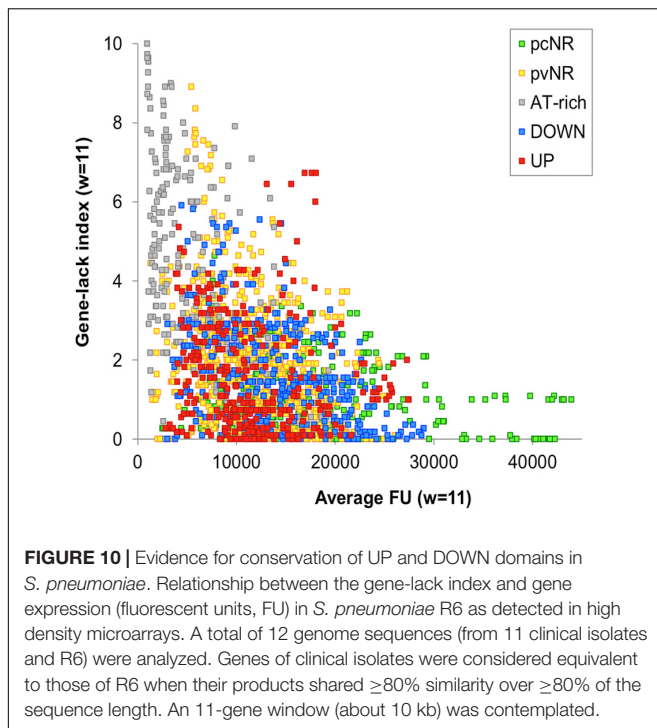
### Importance of the Protein Interaction Network

A significant fraction of the pcNR genes codes for proteins with important roles in central metabolism and that have a high number of protein-protein interactions (PPIs). PPIs provide a rough estimate of a protein's importance in cell physiology. The estimated amounts of protein produced, and their functions, support the idea that the genes of pcNR domains are more involved in the central metabolic network than are those of the pvNR domains. In stark contrast, ATr genes appear to play little or no role in central metabolism; their

PPI values are at most only about one third of the average for the remaining genome. As mentioned above, changes in the location of genes could lead to alterations in cell physiology, which holds true for both central metabolic (Soler-Bistue et al., 2015) and regulatory genes (Gerganova et al., 2015). The physical positioning of specific supercoiling-favorable regions in the chromosome is also related to the ability to gain access to cytoplasmic regions rich in ribosomes (Soler-Bistue et al., 2015).

Overall, the evidence supports the idea that the function, expression, essentiality and stability of genomic positions are interconnected, as reported for *Dickeya dadantii* and *E. coli* (Sobetzko et al., 2012; Jiang et al., 2015). Altogether, the pcNR genes reflect a multistep adaptation in the transcription-translation-interaction cascade that facilitates the activity of these genes' products, thereby increasing bacterial fitness.





### Pathogenesis and Immunogenicity

DNA topology regulates the expression of virulence factors in several bacteria (Dorman and Porter, 1998; Cameron and Dorman, 2012; Reverchon and Nasser, 2013; Jiang et al., 2015). In *S. pneumoniae*, three types of virulence genes show differences in their distribution among domains. Widely accepted virulence factors are more abundant in pvNR domains (**Figure 12B**), while genes contributing [as estimated by signature-tagged mutagenesis (Hensel et al., 1995)] to intranasal colonization, meningitis or otitis (Chen et al., 2008; Molzen et al., 2011) are more abundant in DOWN domains (**Figure 12C**). Finally, genes coding for proteins that trigger an immune response in humans (Giefing et al., 2009), and which are therefore candidate targets for a serotype-independent protein-based vaccine against pneumococcus, are predominant in the pvNR domains. The pvNR domains also contain more genes coding for extracellular proteins or proteins anchored in the cell wall than do pcNR domains. All in all, pvNR domains show strong allelic variation by being subjected to selective pressure during adhesion, cytotoxic challenge and immune system evasion. This variation also increases the genome pool of the species via gene duplication/paralogs in which one copy is not subject to immediate pressure (Mira et al., 2010). The link between supercoiling stress and virulence enhancement does not seem to be the rule for *S. pneumoniae*, the canonical virulence and accessory factors of which are preferentially encoded in the pvNR or DOWN domains.

### Genes Involved in Competence

Gene transfer is a primary driver of evolution in bacteria, but the introduction of new genetic material at random can

perturb chromosomal topology. *S. pneumoniae* is a naturally transformable bacterium (Claverys et al., 2006; Martin et al., 2006), the evolution of which (including its antibiotic-resistance and virulence factors) depends on both intra-species and inter-species chromosomal transformation (Dowson et al., 1989, 1990; Balsalobre et al., 2003; Ferrándiz et al., 2005). Competence involves the transient transcriptional modulation of  $\sim 10\%$  of the genome with strict timing (Peterson et al., 2004). When under stress (the X state), the competence system -which bears some resemblance to the SOS repair system of *E. coli* and other bacteria - is activated (Claverys et al., 2006). In fact, FQs induce the SOS response since they cause double-strand breaks in chromosomes (Drlica et al., 2008). As described above, local supercoiling changes triggered by FQs activate pneumococcal competence, but global supercoiling changes do so too. The early and delayed up-regulated competence genes (those activated during stress) are mainly located in UP domains. Many pcNR genes are, however, down-regulated, indicating that during the X-state the topology of the chromosome is perturbed to a degree that threatens cell viability via effects on the central metabolic machinery. This explains why growth is slowed during competence (Oggioni et al., 2004) and why several mechanisms have been acquired, including the use of small untranslated RNAs and proteases to actively terminate the X-state and promptly recover the normal topological situation (Echenique et al., 2000; Cassone et al., 2012).

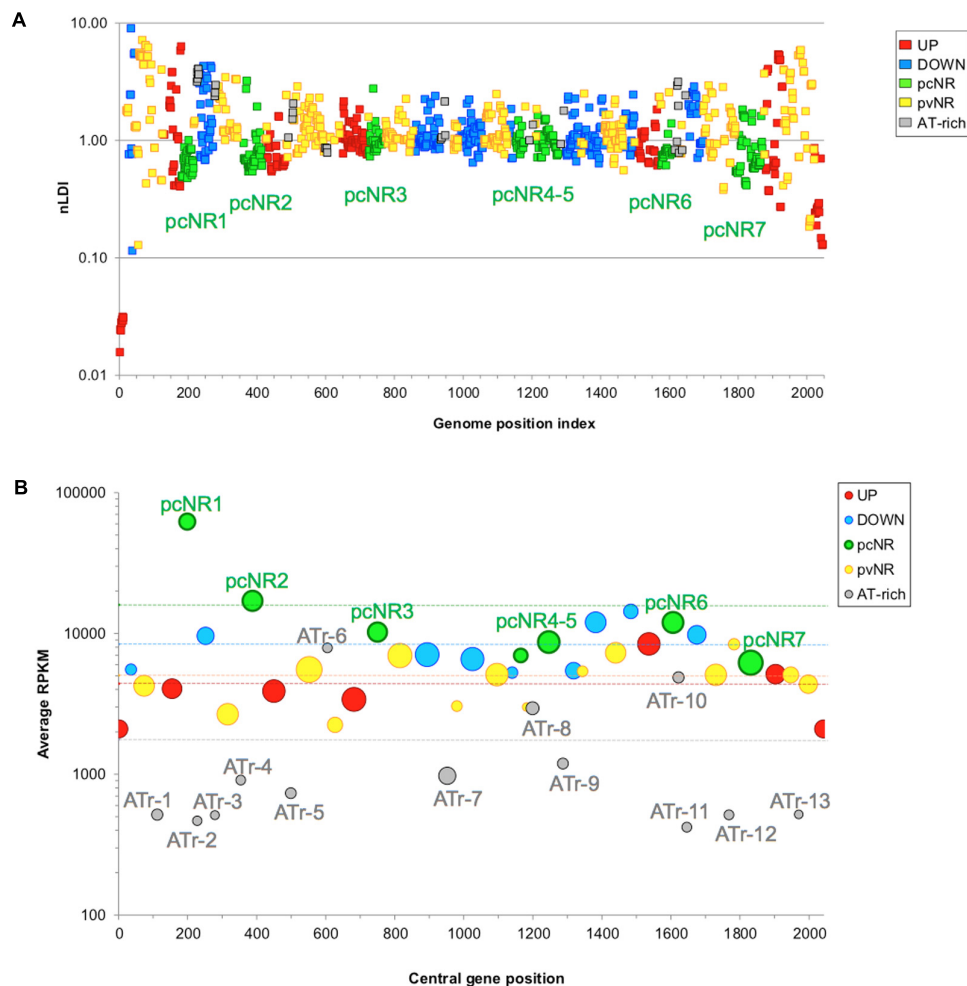
### Horizontally Acquired Genes

In *S. pneumoniae* R6, up to 12.1% of the genome is thought to have been acquired by horizontal gene transfer. The distribution of these acquired genes among domains is uneven, with a clear bias toward ATr domains (**Figure 12D**). This suggests that these domains act as structural or parasitic DNA hotspots, which agrees with their low transcriptional level and annotated functions (Ferrándiz et al., 2010, 2014, 2016a). It remains open the possibility that the ATr regions influence the organization of topological dynamics, or that they are involved in the acquisition of foreign genes.

## CONCLUSION AND PERSPECTIVES

The transcriptome of *S. pneumoniae* alters with local or global changes in supercoiling. Local changes induced by the clinically used FQs LVX, and MOX, which target GyrA and/or Topo IV, trigger a transcriptional response. Both FQs up-regulate the competence regulon in response to stress, and, respectively, cause an increase in intracellular ROS by increasing the uptake of iron (through up-regulation of the *fatDCEB* transporter) and hydrogen peroxide (through up-regulation of the glycolytic pathway), both of which are involved in the Fenton reaction.

Changes in global supercoiling induced by NOV (which targets GyrB), or by SCN (which targets Topo I), have revealed the existence of topological domains that react in a coordinated fashion. In *S. pneumoniae*, the control of DNA-supercoiling occurs mainly via the regulation of transcription of the topoisomerase genes: relaxation triggers the up-regulation of

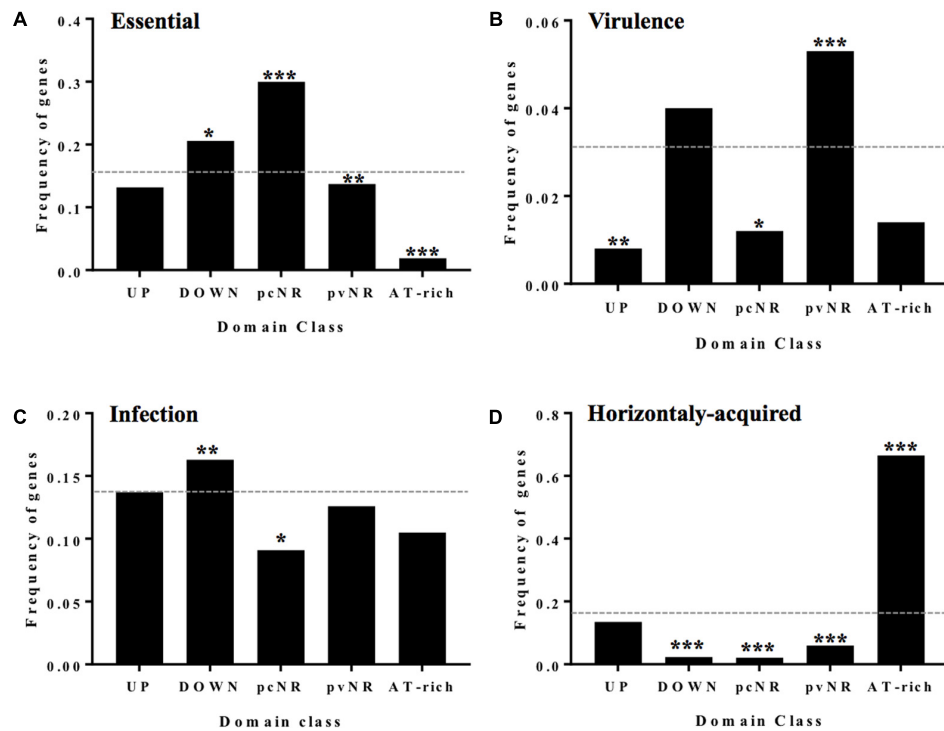


**FIGURE 11 |** Different domains have different locations and transcription levels. **(A)** Location of NOV-reactive and pcNR domains. Dots indicate nLDI values for *S. pneumoniae* R6 genes for the whole *Streptococcus* genus. A total of 1216 genes coding for proteins in *S. pneumoniae* R6, with homologs in  $\geq 10$  species, were contemplated in the analysis. Gene indexing starts from the replication origin. **(B)** Abundance of RNA transcripts according to domain type. RNA-Seq data were normalized by gene size (kb) using the reads per kb per million mapped reads (RPKM). The positions of the domains in the genome are represented by the position of the central gene (central gene position, X-axis). The size of the circles is proportional to the number of genes in the domain. Only pvNR clusters with at least 10 members are shown. Horizontal lines in the corresponding color indicate the average RPKM value for each domain class. Taken from Martín-Galiano et al. (2017).

*gyrA* and *gyrB* and the down-regulation of the Topo I (*topA*) and Topo IV (*parEC*) genes, while hypernegative supercoiling triggers the down-regulation of *topA*. The transcription of *gyrB* and *topA* is regulated by their strategic chromosomal location in the topological domains, while the expression of *parEC* and *gyrA* depends on the specific regulation of their promoters. Although the regulators of *parEC* are unknown, the promoter of *gyrA* shows an intrinsic curvature that acts as a sensor of the supercoiling level. In addition, chromatin immunoprecipitation experiments have revealed Topo I to bind to the *gyrA* promoter. Therefore, Topo I, the transcription of which is regulated by the supercoiling level, appears to regulate *gyrA* expression.

The regulation of topoisomerase genes is part of a global response to changes in supercoiling. Relaxation affects  $> 13\%$  of the genome (from 13 to 24%), while hypernegative supercoiling affects 10%. In both cases, responsive genes are grouped into

domains that essentially overlap, suggesting that they have a fixed chromosomal location. Based on their structural and functional characteristics, and the change in the domains detected under relaxation, the following types can be defined: UP, DOWN, pcNR, pvNR, and ATr. The genes of the UP, DOWN, and pcNR domains have been found at equivalent positions present in most streptococci, especially near the Ori. pcNR domains are interleaved between UP, DOWN, and pvNR domains, which suggests a higher-order macrostructural unit. The pcNRs genes show the highest level of transcription, and contain most of the essential genes plus those involved in the central metabolic network. In stark contrast, the ATr domains show the lowest transcriptional levels, and the genes they contain appear to have little to do with the central metabolic network. This explains the tropism of pcNR genes for topologically secure areas, helping to maintain the constant provision of central proteins.



**FIGURE 12 |** Genes of the different domains have different functionalities. **(A)** Fraction of essential genes in different domains. **(B)** Virulence factors. **(C)** Horizontally acquired genes abundance and domain class. **(D)** Essential genes for infection by STM. Statistical significance with respect to the genome average (dashed line): \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ . Taken from Martín-Galiano et al. (2017), with modifications.

The genes coding for the classical virulence factors, plus those coding for immunogenic proteins, are more common in the pvNR domains, while genes contributing toward the establishment of infection are more common in the DOWN domains. The distribution of horizontally acquired genes is clearly biased toward ATr domains, suggesting these to be hotspots for the acquisition of foreign genes.

In general, UP gene expression is favored by topological stress; DOWN genes are highly expressed under favorable conditions and less so during such stress. ATr domains may sense topological stress and modify supercoiling in their area to reduce the transcription of adjacent genes, preferentially those in the DOWN domains. The chromosome supercoiling structure may act as a multi-sensor with homeostatic capacity, adapted to react to unfavorable conditions.

Pneumococcal genes appear to be subject to topology-driven selection that defines the chromosomal location of genes involved in metabolism, virulence and competence. Together, these organizational features reveal the genome of *S. pneumoniae* to be influenced by physiology-related topological rules. A global topology theory might be envisaged in which gene positioning is far from random. Many aspects of the importance of gene location – such as the idiosyncrasy of the domains and how this affects fundamental aspects of bacterial biology – are only now becoming understood.

Topological genomics – topogenomics – provides an alternative paradigm of genome analysis. Certainly, genome

architecture plays an important role in the pathobiology and evolution of *S. pneumoniae*, and it is tempting to speculate that in other species too, the genes are subjected to topology-driven selection pressure that defines their chromosomal locations. Data from many species will, however, be needed before a full understanding of all the rules underlying topogenomics are known and understood.

## AUTHOR CONTRIBUTIONS

All authors made intellectual contributions to the work and approved it for publication. AdC supervised all the studies and wrote the manuscript. MF performed most of the experiments related to determinations of supercoiling densities and transcriptomic studies. AM-G performed the bioinformatic studies. MG performed the characterization of topoisomerase I and its inhibition by seconeolitsine. JT-V contributed to the experiments of chromatin immunoprecipitation.

## FUNDING

AM-G is funded by a Miguel Servet contract from the Instituto de Salud Carlos III-MINECO. This work was supported by the Ministerio de Economía y Competitividad (BIO2014-55462-R).

## ACKNOWLEDGMENTS

We thank Monica Amblar (*Centro Nacional de Microbiología, ISCIII, Madrid, Spain*), Pablo Hernández (*Centro de*

*Investigaciones Biológicas, CSIC, Madrid, Spain*), and Pedro A. Lazo-Zbikowski (*Instituto de Biología Molecular y Celular del Cáncer, CSIC, Salamanca, Spain*), and for their critical reading of the manuscript.

## REFERENCES

- Arnoldi, E., Pan, X. S., and Fisher, L. M. (2013). Functional determinants of gate-DNA selection and cleavage by bacterial type II topoisomerases. *Nucleic Acids Res.* 41, 9411–9423. doi: 10.1093/nar/gkt696
- Badrinarayanan, A., Le, T. B., and Laub, M. T. (2015). Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.* 31, 171–199. doi: 10.1146/annurev-cellbio-100814-125211
- Balas, D., Fernández-Moreira, E., and de la Campa, A. G. (1998). Molecular characterization of the gene encoding the DNA gyrase A subunit of *Streptococcus pneumoniae*. *J. Bacteriol.* 180, 2854–2861.
- Balsalobre, L., and de la Campa, A. G. (2008). Fitness of *Streptococcus pneumoniae* fluoroquinolone-resistant strains with topoisomerase IV recombinant genes. *Antimicrob. Agents Chemother.* 52, 822–830. doi: 10.1128/AAC.00731-07
- Balsalobre, L., Ferrándiz, M. J., Liñares, J., Tubau, F., and de la Campa, A. G. (2003). Viridans group streptococci are donors in horizontal transfer of topoisomerase IV genes to *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 47, 2072–2081. doi: 10.1128/AAC.47.7.2072-2081.2003
- Blot, N., Mavathur, R., Geertz, M., Travers, A., and Muskhelishvili, G. (2006). Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.* 7, 710–715. doi: 10.1038/sj.embor.7400729
- Brito, L., Wilton, J., Ferrándiz, M. J., Gómez-Sanz, A., de la Campa, A. G., and Amblar, M. (2017). Absence of tmRNA has a protective effect against fluoroquinolones in *Streptococcus pneumoniae*. *Front. Microbiol.* 7:2164. doi: 10.3389/fmicb.2016.02164
- Browning, D. F., and Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2, 57–65. doi: 10.1038/nrmicro787
- Cameron, A. D., and Dorman, C. J. (2012). A fundamental regulatory mechanism operating through OmpR and DNA topology controls expression of *Salmonella* pathogenicity islands SPI-1 and SPI-2. *PLoS Genet.* 8:e1002615. doi: 10.1371/journal.pgen.1002615
- Cassone, M., Gagne, A. L., Spruce, L. A., Seeholzer, S. H., and Seibert, M. E. (2012). The HtrA protease from *Streptococcus pneumoniae* digests both denatured proteins and the competence-stimulating peptide. *J. Biol. Chem.* 287, 38449–38459. doi: 10.1074/jbc.M112.391482
- Champoux, J. J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* 70, 369–413. doi: 10.1146/annurev.biochem.70.1.369
- Chandler, M. G., and Pritchard, R. H. (1975). The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. *Mol. Gen. Genet.* 138, 127–141. doi: 10.1007/BF02428117
- Chen, H., Ma, Y., Yang, J., O'Brien, C. J., Lee, S. L., Mazurkiewicz, J. E., et al. (2008). Genetic requirement for pneumococcal ear infection. *PLoS ONE* 3:e2950. doi: 10.1371/journal.pone.0002950
- Cheng, B., Liu, I. F., and Tse-Dinh, Y. C. (2007). Compounds with antibacterial activity that enhance DNA cleavage by bacterial DNA topoisomerase I. *J. Antimicrob. Chemother.* 59, 640–645. doi: 10.1093/jac/dkl556
- Cheng, B., Zhu, C. X., Ji, C., Ahumada, A., and Tse-Dinh, Y. C. (2003). Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *J. Biol. Chem.* 278, 30705–30710. doi: 10.1074/jbc.M303403200
- Claret, L., and Rouviere-Yaniv, J. (1996). Regulation of HU alpha and HU beta by CRP and FIS in *Escherichia coli*. *J. Mol. Biol.* 263, 126–139. doi: 10.1006/jmbi.1996.0564
- Claverys, J. P., Prudhomme, M., and Martin, B. (2006). Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu. Rev. Microbiol.* 60, 451–475. doi: 10.1146/annurev.micro.60.080805.142139
- Croucher, N. J., Vernikos, G. S., Parkhill, J., and Bentley, S. D. (2011). Identification, variation and transcription of pneumococcal repeat sequences. *BMC Genomics* 12:120. doi: 10.1186/1471-2164-12-120
- Crozat, E., Philippe, N., Lenski, R. E., Geiselman, J., and Schneider, D. (2005). Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169, 523–532. doi: 10.1534/genetics.104.035717
- Deng, S., Stein, R. A., and Higgins, N. P. (2005). Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.* 57, 1511–1521. doi: 10.1111/j.1365-2958.2005.04796.x
- Dillon, S. C., and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* 8, 185–195. doi: 10.1038/nrmicro2261
- DiNardo, S., Voelkel, K. A., Sternglanz, R., Reynolds, A. E., and Wright, A. (1982). *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. *Cell* 31, 43–51. doi: 10.1016/0092-8674(82)90403-2
- Dorman, C. J. (2013). Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.* 11, 349–355. doi: 10.1038/nrmicro3007
- Dorman, C. J., and Porter, M. E. (1998). The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. *Mol. Microbiol.* 29, 677–684. doi: 10.1046/j.1365-2958.1998.00902.x
- Dowson, C. G., Hutchison, A., Brannigan, J. A., George, R. C., Hansman, D., Liñares, J., et al. (1989). Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* 86, 8842–8846. doi: 10.1073/pnas.86.22.8842
- Dowson, C. G., Hutchinson, A., Woodford, N., Johnson, A. P., George, R. C., and Spratt, B. G. (1990). Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* 87, 5858–5862. doi: 10.1073/pnas.87.15.5858
- Drlica, K. (1992). Control of bacterial DNA supercoiling. *Mol. Microbiol.* 6, 425–433. doi: 10.1111/j.1365-2958.1992.tb01486.x
- Drlica, K., Malik, M., Kerns, R. J., and Zhao, X. (2008). Quinolone-mediated bacterial death. *Antimicrob. Agents Chemother.* 52, 385–392. doi: 10.1128/AAC.01617-06
- Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.* 59, 723–730. doi: 10.1111/j.1365-2958.2005.05006.x
- Drolet, M., Bi, X., and Liu, L. F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J. Biol. Chem.* 269, 2068–2074.
- Dupaigne, P., Tonthat, N. K., Espeli, O., Whitfill, T., Boccard, F., and Schumacher, M. A. (2012). Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chromosome. *Mol. Cell* 48, 560–571. doi: 10.1016/j.molcel.2012.09.009
- Dwyer, D. J., Belenky, P. A., Yang, J. H., MacDonald, I. C., Martell, J. D., Takahashi, N., et al. (2014). Antibiotics induce redox-related physiological alterations as part of their lethality. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2100–E2109. doi: 10.1073/pnas.1401876111
- Dwyer, D. J., Collins, J. J., and Walker, G. C. (2015). Unraveling the physiological complexities of antibiotic lethality. *Annu. Rev. Pharmacol. Toxicol.* 55, 313–332. doi: 10.1146/annurev-pharmtox-010814-124712
- Dwyer, D. J., Kohanski, M. A., Hayete, B., and Collins, J. J. (2007). Gyrase inhibitors induce an oxidative damage cellular death pathway in *Escherichia coli*. *Mol. Syst. Biol.* 3, 91. doi: 10.1038/msb4100135
- Echenique, J. R., Chapuy-Regaud, S., and Trombe, M. C. (2000). Competence regulation by oxygen in *Streptococcus pneumoniae*: involvement of ciaRH and comCDE. *Mol. Microbiol.* 36, 688–696. doi: 10.1046/j.1365-2958.2000.01891.x
- Espeli, O., Mercier, R., and Boccard, F. (2008). DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.* 68, 1418–1427. doi: 10.1111/j.1365-2958.2008.06239.x
- Falconi, M., Brandi, A., La Teana, A., Gualerzi, C. O., and Pon, C. L. (1996). Antagonistic involvement of FIS and H-NS proteins in the transcriptional control of hns expression. *Mol. Microbiol.* 19, 965–975. doi: 10.1046/j.1365-2958.1996.436961.x



- Fang, G., Rocha, E., and Danchin, A. (2005). How essential are nonessential genes? *Mol. Biol. Evol.* 22, 2147–2156. doi: 10.1093/molbev/msi211
- Fang, G., Rocha, E. P., and Danchin, A. (2008). Persistence drives gene clustering in bacterial genomes. *BMC Genomics* 9:4. doi: 10.1186/1471-2164-9-4
- Fenoll, A., Gimenez, M. J., Vicioso, M. D., Granizo, J. J., Robledo, O., and Aguilar, L. (2009). Susceptibility of pneumococci causing meningitis in Spain and prevalence among such isolates of serotypes contained in the 7-valent pneumococcal conjugate vaccine. *J. Antimicrob. Chemother.* 64, 1338–1340. doi: 10.1093/jac/dkp376
- Fernández-Moreira, E., Balas, D., González, I., and de la Campa, A. G. (2000). Fluoroquinolones inhibit preferentially *Streptococcus pneumoniae* DNA topoisomerase IV than DNA gyrase native proteins. *Microb. Drug Resist.* 6, 259–267. doi: 10.1089/mdr.2000.6.259
- Ferrándiz, M. J., Ardanuy, C., Liñares, J., García-Arenzana, J. M., Cercenado, E., Fleites, A., et al. (2005). New mutations and horizontal transfer of *rpoB* among rifampin-resistant *Streptococcus pneumoniae* from four Spanish hospitals. *Antimicrob. Agents Chemother.* 49, 2237–2245. doi: 10.1128/AAC.49.6.2237-2245.2005
- Ferrándiz, M. J., Arnanz, C., Martín-Galiano, A. J., Rodríguez-Martin, C., and de la Campa, A. G. (2014). Role of global and local topology in the regulation of gene expression in *Streptococcus pneumoniae*. *PLoS ONE* 9:e101574. doi: 10.1371/journal.pone.0101574
- Ferrándiz, M. J., and de la Campa, A. G. (2014). The fluoroquinolone levofloxacin triggers the transcriptional activation of iron transport genes that contribute to cell death in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 58, 247–257. doi: 10.1128/AAC.01706-13
- Ferrándiz, M. J., Martín-Galiano, A. J., Arnanz, C., Camacho-Soguero, I., Tirado-Vélez, J. M., and de la Campa, A. G. (2016a). An increase in negative supercoiling in bacteria reveals topology-reacting gene clusters and a homeostatic response mediated by the DNA topoisomerase I gene. *Nucleic Acids Res.* 44, 7292–7303. doi: 10.1093/nar/gkw602
- Ferrándiz, M. J., Martín-Galiano, A. J., Arnanz, C., Zimmerman, T., and de la Campa, A. G. (2016b). Reactive oxygen species contribute to the bactericidal effects of the fluoroquinolone moxifloxacin in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 60, 409–417. doi: 10.1128/AAC.02299-15
- Ferrándiz, M. J., Martín-Galiano, A. J., Schvartzman, J. B., and de la Campa, A. G. (2010). The genome of *Streptococcus pneumoniae* is organized in topology-reacting gene clusters. *Nucleic Acids Res.* 38, 3570–3581. doi: 10.1093/nar/gkq106
- French, S. (1992). Consequences of replication fork movement through transcription units in vivo. *Science* 258, 1362–1365. doi: 10.1126/science.1455232
- García, M. T., Blázquez, M. A., Ferrándiz, M. J., Sanz, M. J., Silva-Martín, N., Hermoso, J. A., et al. (2011). New alkaloid antibiotics that target the DNA topoisomerase I of *Streptococcus pneumoniae*. *J. Biol. Chem.* 286, 6402–6413. doi: 10.1074/jbc.M110.148148
- Gellert, M., Mizuuchi, K., O'Dea, H., and Nash, H. A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3872–3876. doi: 10.1073/pnas.73.11.3872
- Gerganova, V., Berger, M., Zaldastanishvili, E., Sobetzko, P., Lafon, C., Mourez, M., et al. (2015). Chromosomal position shift of a regulatory gene alters the bacterial phenotype. *Nucleic Acids Res.* 43, 8215–8226. doi: 10.1093/nar/gkv709
- Giefing, C., Nagy, E., and von Gabain, A. (2009). The antigenome: from protein subunit vaccines to antibody treatments of bacterial infections? *Adv. Exp. Med. Biol.* 655, 90–117. doi: 10.1007/978-1-4419-1132-2\_9
- Gmüender, H., Kuratli, K., Di Padova, K., Gray, C. P., Keck, W., and Evers, S. (2001). Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res.* 11, 28–42. doi: 10.1101/gr.157701
- Grainger, D. C., Goldberg, M. D., Lee, D. J., and Busby, S. J. (2008). Selective repression by Fis and H-NS at the *Escherichia coli* *dps* promoter. *Mol. Microbiol.* 68, 1366–1377. doi: 10.1111/j.1365-2958.2008.06253.x
- Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., and Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269, 400–403. doi: 10.1126/science.7618105
- Higgins, N. P., Yang, X., Fu, Q., and Roth, J. R. (1996). Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J. Bacteriol.* 178, 2825–2835. doi: 10.1128/jb.178.10.2825-2835.1996
- Holmes, V. F., and Cozzarelli, N. R. (2000). Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1322–1324. doi: 10.1073/pnas.040576797
- Jacobs, M. R., Felmingham, D., Appelbaum, P. C., Grüneberg, R. N., and Group, T. A. P. (2003). The Alexander project 1998–200: susceptibility of pathogens isolated from community-acquired respiratory tract infection to commonly used antimicrobial agents. *J. Antimicrob. Chemother.* 52, 229–246. doi: 10.1093/jac/dkg321
- Janoir, C., Zeller, V., Kitzis, M.-D., Moreau, N. J., and Gutmann, L. (1996). High-level fluoroquinolone resistance in *Streptococcus pneumoniae* requires mutations in *parC* and *gyrA*. *Antimicrob. Agents Chemother.* 40, 2760–2764.
- Jiang, X., Sobetzko, P., Nasser, W., Reverchon, S., and Muskhelishvili, G. (2015). Chromosomal “stress-response” domains govern the spatiotemporal expression of the bacterial virulence program. *MBio* 6, e00353–15. doi: 10.1128/mBio.00353-15
- Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., and Suzuki, H. (1990). New topoisomerase essential for chromosome segregation in *E. coli*. *Cell* 63, 393–404. doi: 10.1016/0092-8674(90)90172-B
- Kavenoff, R., and Bowen, B. C. (1976). Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. *Chromosoma* 59, 89–101. doi: 10.1007/BF00328479
- Khedkar, S., and Seshasayee, A. S. (2016). Comparative genomics of interreplicore translocations in bacteria: a measure of chromosome topology? *G3 (Bethesda)* 6, 1597–1606. doi: 10.1534/g3.116.028274
- Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A., and Collins, J. J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* 130, 797–810. doi: 10.1016/j.cell.2007.06.049
- Kyaw, M. H., Lynfield, R., Schaffner, W., Craig, A. S., Hadler, J., Reingold, A., et al. (2006). Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant *Streptococcus pneumoniae*. *N. Engl. J. Med.* 354, 1455–1463. doi: 10.1056/NEJMoa051642
- Lathe, W. C. III, Snel, B., and Bork, P. (2000). Gene context conservation of a higher order than operons. *Trends Biochem. Sci.* 25, 474–479. doi: 10.1016/S0968-0004(00)01663-7
- Le, T. B., Imakaev, M. V., Mirny, L. A., and Laub, M. T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* 342, 731–734. doi: 10.1126/science.1242059
- Lefebvre, T., and Stanhope, M. J. (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* 8:R71. doi: 10.1186/gb-2007-8-5-r71
- Leo, E., Gould, K. A., Pan, X. S., Capranico, G., Sanderson, M. R., Palumbo, M., et al. (2005). Novel symmetric and asymmetric DNA scission determinants for *Streptococcus pneumoniae* topoisomerase IV and gyrase are clustered at the DNA breakage site. *J. Biol. Chem.* 280, 14252–14263. doi: 10.1074/jbc.M500156200
- Liñares, J., Ardanuy, C., Pallarés, R., and Fenoll, A. (2010). Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. *Clin. Microbiol. Infect.* 16, 402–410. doi: 10.1111/j.1469-0691.2010.03182.x
- Marr, C., Geertz, M., Hutt, M. T., and Muskhelishvili, G. (2008). Dissecting the logical types of network control in gene expression profiles. *BMC Syst. Biol.* 2:18. doi: 10.1186/1752-0509-2-18
- Martin, B., Quentin, Y., Fichant, G., and Claverys, J. P. (2006). Independent evolution of competence regulatory cascades in streptococci? *Trends Microbiol.* 14, 339–345. doi: 10.1016/j.tim.2006.06.007
- Martín-Galiano, A. J., Ferrándiz, M. J., and de la Campa, A. G. (2017). Bridging chromosomal architecture and pathophysiology of *Streptococcus pneumoniae*. *Genome Biol. Evol.* 9, 350–361. doi: 10.1093/gbe/evw299
- Martín-Galiano, A. J., Wells, J. M., and de la Campa, A. G. (2004). Relationship between codon biased genes, microarray expression values and physiological characteristics of *Streptococcus pneumoniae*. *Microbiology* 150, 2313–2325. doi: 10.1099/mic.0.27097-0
- Masse, E., and Drolet, M. (1999). *Escherichia coli* DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J. Biol. Chem.* 274, 16659–16664. doi: 10.1074/jbc.274.23.16659

- Menzel, R., and Gellert, M. (1983). Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell* 34, 105–113. doi: 10.1016/0092-8674(83)90140-X
- Menzel, R., and Gellert, M. (1987a). Fusions of the *Escherichia coli* *gyrA* and *gyrB* control regions to the galactokinase gene are inducible by coumermycin treatment. *J. Bacteriol.* 169, 1272–1278.
- Menzel, R., and Gellert, M. (1987b). Modulation of transcription by DNA supercoiling: a deletion analysis of the *Escherichia coli* *gyrA* and *gyrB* promoters. *Proc. Natl. Acad. Sci. U.S.A.* 84, 4185–4189.
- Mira, A., Martín-Cuadrado, A. B., D'Auria, G., and Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int. Microbiol.* 13, 45–57. doi: 10.2436/20.1501.01.110
- Molzen, T. E., Burghout, P., Bootsma, H. J., Brandt, C. T., van, der Gaast-de Jongh, C. E., Eleveld, M. J., et al. (2011). Genome-wide identification of *Streptococcus pneumoniae* genes essential for bacterial replication during experimental meningitis. *Infect. Immun.* 79, 288–297. doi: 10.1128/IAI.00631-10
- Moore, M. R., Gertz, J. R. E., Woodbury, R. L., Barkocy-Gallagher, G. A., Schaffner, W., Lexau, C., et al. (2008). Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. *J. Infect. Dis.* 197, 1016–1027. doi: 10.1086/528996
- Muñoz, R., and de la Campa, A. G. (1996). ParC subunit of DNA topoisomerase IV of *Streptococcus pneumoniae* is a primary target of fluoroquinolones and cooperates with DNA gyrase A subunit in forming resistance phenotype. *Antimicrob. Agents Chemother.* 40, 2252–2257.
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500
- Ogawa, T., Yogo, K., Furukie, S., Sutoh, K., Kikuchi, A., and Kinoshita, K. Jr. (2015). Direct observation of DNA overwinding by reverse gyrase. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7495–7500. doi: 10.1073/pnas.1422203112
- Oggioni, M. R., Iannelli, F., Ricci, S., Chiavolini, D., Parigi, R., Trappetti, C., et al. (2004). Antibacterial activity of a competence-stimulating peptide in experimental sepsis caused by *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 48, 4725–4732. doi: 10.1128/AAC.48.12.4725-4732.2004
- Pérez-Martín, J., and Espinosa, M. (1991). The RepA repressor can act as a transcriptional activator by inducing DNA bends. *EMBO J.* 10, 1375–1382.
- Pérez-Martín, J., Rojo, F., and de Lorenzo, V. (1994). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol. Rev.* 58, 268–290.
- Peter, B. J., Arsuaga, J., Breier, A. M., Khodursky, A. B., Brown, P. O., and Cozzarelli, N. R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.* 5:R87. doi: 10.1186/gb-2004-5-11-r87
- Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Sniesrud, E. C., Luo, P., et al. (2004). Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol. Microbiol.* 51, 1051–1070. doi: 10.1046/j.1365-2958.2003.03907.x
- Phoenix, P., Raymond, M. A., Masse, E., and Drolet, M. (1997). Roles of DNA topoisomerases in the regulation of R-loop formation in vitro. *J. Biol. Chem.* 272, 1473–1479. doi: 10.1074/jbc.272.3.1473
- Pilishvili, T., Lexau, C., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., et al. (2010). Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J. Infect. Dis.* 201, 32–41. doi: 10.1086/648593
- Postow, L., Crisone, N. J., Peter, B. J., Hardy, C. D., and Cozzarelli, N. R. (2001). Topological challenges to DNA replication: conformations at the fork. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8219–8226. doi: 10.1073/pnas.111006998
- Postow, L., Hardy, C. D., Arsuaga, J., and Cozzarelli, N. R. (2004). Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* 18, 1766–1779. doi: 10.1101/gad.1207504
- Prudhomme, M., Attaiach, L., Sanchez, G., Martin, B., and Claverys, J. P. (2006). Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae*. *Science* 313, 89–92. doi: 10.1126/science.1127912
- Pruss, G. J., Manes, S. H., and Drlica, K. (1982). *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell* 31, 35–42. doi: 10.1016/0092-8674(82)90402-0
- Reams, A. B., and Neidle, E. L. (2004). Selection for gene clustering by tandem duplication. *Annu. Rev. Microbiol.* 58, 119–142. doi: 10.1146/annurev.micro.58.030603.123806
- Reverchon, S., and Nasser, W. (2013). Dickeya ecology, environment sensing and regulation of virulence programme. *Environ. Microbiol. Rep.* 5, 622–636. doi: 10.1111/1758-2229.12073
- Rocha, E. P. (2004a). Order and disorder in bacterial genomes. *Curr. Opin. Microbiol.* 7, 519–527. doi: 10.1016/j.mib.2004.08.006
- Rocha, E. P. (2004b). The replication-related organization of bacterial genomes. *Microbiology* 150, 1609–1627. doi: 10.1099/mic.0.26974-0
- Rocha, E. P., and Danchin, A. (2003). Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat. Genet.* 34, 377–378. doi: 10.1038/ng1209
- Schmid, M. B., and Roth, J. R. (1987). Gene location affects expression level in *Salmonella typhimurium*. *J. Bacteriol.* 169, 2872–2875. doi: 10.1128/jb.169.6.2872-2875.1987
- Schneider, R., Travers, A., Kutateladze, T., and Muskhelishvili, G. (1999). A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.* 34, 953–964. doi: 10.1046/j.1365-2958.1999.01656.x
- Schroder, W., Bernhardt, J., Marincola, G., Klein-Hitpass, L., Herbig, A., Krupp, G., et al. (2014). Altering gene expression by aminocoumarins: the role of DNA supercoiling in *Staphylococcus aureus*. *BMC Genomics* 15:291. doi: 10.1186/1471-2164-15-291
- Sinden, R. R., and Pettijohn, D. E. (1981). Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc. Natl. Acad. Sci. U.S.A.* 78, 224–228. doi: 10.1073/pnas.78.1.224
- Slager, J., and Veening, J. W. (2016). Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol.* 24, 788–800. doi: 10.1016/j.tim.2016.06.003
- Snyder, M., and Drlica, K. (1979). DNA gyrase on the bacterial chromosome: DNA cleavage induced by oxolinic acid. *J. Mol. Biol.* 131, 287–302. doi: 10.1016/0022-2836(79)90077-9
- Sobetzko, P., Travers, A., and Muskhelishvili, G. (2012). Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc. Natl. Acad. Sci. U.S.A.* 109, E42–E50. doi: 10.1073/pnas.1108229109
- Soler-Bistue, A., Mondotte, J. A., Bland, M. J., Val, M. E., Saleh, M. C., and Mazel, D. (2015). Genomic location of the major ribosomal protein gene locus determines *Vibrio cholerae* global growth and infectivity. *PLoS Genet.* 11:e1005156. doi: 10.1371/journal.pgen.1005156
- Stassi, D., López, P., Espinosa, M., and Lacks, S. A. (1981). Cloning of chromosomal genes in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* 78, 7028–7032. doi: 10.1073/pnas.78.11.7028
- Tankovic, J., Perichon, B., Duval, J., and Courvalin, P. (1996). Contribution of mutations in *gyrA* and *parC* genes to fluoroquinolone resistance of mutants of *Streptococcus pneumoniae* obtained in vivo and in vitro. *Antimicrob. Agents Chemother.* 40, 2505–2510.
- Thiara, A. S., and Cundliffe, E. (1989). Interplay of novobiocin-resistant and -sensitive DNA gyrase activities in selfprotection of the novobiocin producer *Streptomyces spheeroides*. *Gene* 81, 65–72. doi: 10.1016/0378-1119(89)90337-5
- Travers, A., and Muskhelishvili, G. (2005). DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.* 3, 157–169. doi: 10.1038/nrmicro1088
- Tse-Dinh, Y.-C. (1985). Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. *Nucleic Acids Res.* 13, 4751–4763. doi: 10.1093/nar/13.13.4751
- Unniraman, S., Chatterji, M., and Nagaraja, V. (2002). DNA gyrase genes in *Mycobacterium tuberculosis*: a single operon driven by multiple promoters. *J. Bacteriol.* 184, 5449–5456. doi: 10.1128/JB.184.19.5449-5456.2002
- van Opijnen, T., and Camilli, A. (2012). A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res.* 22, 2541–2551. doi: 10.1101/gr.137430.112
- Vora, T., Hottes, A. K., and Tavazoie, S. (2009). Protein occupancy landscape of a bacterial genome. *Mol. Cell* 35, 247–253. doi: 10.1016/j.molcel.2009.06.035
- Wang, X., Montero Llopis, P., and Rudner, D. Z. (2013). Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.* 14, 191–203. doi: 10.1038/nrg3375
- Wang, X., and Zhao, X. (2009). Contribution of oxidative damage to antimicrobial lethality. *Antimicrob. Agents Chemother.* 53, 1395–1402. doi: 10.1128/AAC.01087-08

- Wang, X., Zhao, X., Malik, M., and Drlica, K. (2010). Contribution of reactive oxygen species to pathways of quinolone-mediated bacterial cell death. *J. Antimicrob. Chemother.* 65, 520–524. doi: 10.1093/jac/dkp486
- Weinstein-Fischer, D., and Altuvia, S. (2007). Differential regulation of *Escherichia coli* topoisomerase I by Fis. *Mol. Microbiol.* 63, 1131–1144. doi: 10.1111/j.1365-2958.2006.05569.x
- Whitney, C. G., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., Lynfield, R., et al. (2003). Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *N. Engl. J. Med.* 348, 1737–1746. doi: 10.1056/NEJMoa022823
- Worcel, A., and Burgi, E. (1972). On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* 71, 127–147. doi: 10.1016/0022-2836(72)90342-7
- World Health Organization (2007). Pneumococcal conjugate vaccine for childhood immunization-WHO position paper. *Wkly. Epidemiol. Rec.* 82, 93–104.
- Wright, M. A., Kharchenko, P., Church, G. M., and Segre, D. (2007). Chromosomal periodicity of evolutionarily conserved gene pairs. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10559–10564. doi: 10.1073/pnas.0610776104
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 de la Campa, Ferrándiz, Martín-Galiano, García and Tirado-Vélez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Toxin $\zeta$ Triggers a Survival Response to Cope with Stress and Persistence

**María Moreno-del Álamo, Mariangela Tabone<sup>†</sup>, Virginia S. Lioy<sup>†</sup> and Juan C. Alonso<sup>\*</sup>**

Department of Microbial Biotechnology, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Ramon Diaz Orejas,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain  
Nadia Berkova,  
Institut National de la Recherche  
Agronomique (INRA), France

### \*Correspondence:

Juan C. Alonso  
jcalonso@cnb.csic.es

### <sup>†</sup>Present Address:

Mariangela Tabone,  
Department of Basic Biomedical  
Sciences, Faculty of Biomedical  
Sciences and Health, Universidad  
Europea de Madrid, Madrid, Spain;  
Virginia S. Lioy,  
Institute for Integrative Biology of the  
Cell (I2BC), CEA, Centre National de la  
Recherche Scientifique, Univ.  
Paris-Sud, Université Paris-Saclay,  
Gif-sur-Yvette, France.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 17 April 2017

**Accepted:** 02 June 2017

**Published:** 23 June 2017

### Citation:

Moreno-del Álamo M, Tabone M,  
Lioy VS and Alonso JC (2017) Toxin  $\zeta$   
Triggers a Survival Response to Cope  
with Stress and Persistence.  
Front. Microbiol. 8:1130.  
doi: 10.3389/fmicb.2017.01130

Bacteria have evolved complex regulatory controls in response to various environmental stresses. Protein toxins of the  $\zeta$  superfamily, found in prominent human pathogens, are broadly distributed in nature. We show that  $\zeta$  is a uridine diphosphate-N-acetylglucosamine (UNAG)-dependent ATPase whose activity is inhibited *in vitro* by stoichiometric concentrations of  $\epsilon_2$  antitoxin. *In vivo*, transient  $\zeta$  expression promotes a reversible multi-level response by altering the pool of signaling purine nucleotides, which leads to growth arrest (dormancy), although a small cell subpopulation persists rather than tolerating toxin action. High c-di-AMP levels (absence of phosphodiesterase GdpP) decrease, and low c-di-AMP levels (absence of diadenylate cyclase DisA) increase the rate of  $\zeta$  persistence. The absence of CodY, a transition regulator from exponential to stationary phase, sensitizes cells to toxin action, and suppresses persisters formed in the  $\Delta disA$  context. These changes, which do not affect the levels of stochastic ampicillin (Amp) persistence, sensitize cells to toxin and Amp action. Our findings provide an explanation for the connection between  $\zeta$ -mediated growth arrest (with alterations in the GTP and c-di-AMP pools) and persistence formation.

**Keywords:** toxin-antitoxin system, cell wall inhibition, c-di-AMP, CodY, (p)ppGpp, DisA

## INTRODUCTION

The toxin-antitoxin (TA) systems are widely distributed in free-living bacteria, in their extrachromosomal elements, and in archaea (Gerdes, 2013; Unterholzner et al., 2013). The toxins of all known TA systems are proteins while the antitoxins are either proteins or non-coding RNAs. The TA systems are classified into five different TA types (Yamaguchi et al., 2011), being the most broadly distributed the type II TA system, where both the toxin and the antitoxin are proteins (Leplae et al., 2011; Gerdes, 2013). The type II toxins use different strategies to regulate growth control and cellular processes related to the general stress response. Toxins of the  $\zeta$ /PezT superfamily, which are among the most broadly distributed in nature, are found in major human pathogens and in environmentally important bacteria of the phylum Firmicutes (Mutschler and Meinhart, 2013). The plasmid-borne  $\zeta$  gene product from *Streptococcus pyogenes*, *Streptococcus agalactiae* or *Enterococcus faecalis* and the chromosome-encoded  $\zeta$  toxin from *Clostridium perfringens* or *Staphylococcus aureus* (~285 amino acids) share ~43% sequence identity with chromosome-encoded *Streptococcus pneumoniae* or *Streptococcus suis* PezT toxin (~255 amino acids) (reviewed in Mutschler and Meinhart, 2013). When free in solution, these toxins interact with uridine diphosphate-N-acetylglucosamine (UNAG), ATP-Mg<sup>2+</sup> or GTP-Mg<sup>2+</sup> (denoted ATP and GTP), and with their cognate dimeric  $\epsilon$ /PezA antitoxin ( $\epsilon$ /PezA<sub>2</sub>) (Meinhart et al., 2001, 2003; Khoo et al., 2007; Mutschler et al., 2011). A non-toxic heterotetrameric complex



( $\zeta\epsilon_2\zeta$ /PezT-PezA<sub>2</sub>-PezT) interacts with UNAG, but not with ATP/GTP (Meinhart et al., 2001, 2003; Khoo et al., 2007; Mutschler et al., 2011).

Enzymes of the  $\zeta$ /PezT toxin superfamily have a common fold core with phosphotransferases (Meinhart et al., 2001, 2003; Khoo et al., 2007; Mutschler et al., 2011). Toxin  $\zeta$ /PezT transfers the ATP/GTP  $\gamma$ -phosphate to the 3'-hydroxyl group of the UNAG amino sugar, rendering UNAG-3P unreactive and thus reducing cell wall biosynthesis (Mutschler et al., 2011). Although, a quantitative analysis of this reaction showed that in the presence of limiting UNAG and ATP, toxin  $\zeta$  mainly hydrolyzed ATP and only traces of the  $\gamma$ -phosphate are transferred to UNAG (Tabone et al., 2014a).

The fine mechanisms of bacterial responses to toxin action are not generally conserved among different bacterial phyla (Gerdes, 2013). The evolutionary distance between *Escherichia coli* and *Bacillus subtilis*, which is larger than the time divergence between yeasts and humans, reflects the notable differences made by the purine nucleotides in the stringent response (Potrykus and Cashel, 2008; Liu et al., 2015). In *E. coli* (a representative of the  $\gamma$ -proteobacteria class), toxin-mediated persister formation is linked to high levels of guanosine (penta)tetraphosphate ([p]ppGpp), which inhibits the PPX phosphatase; dropping of PPX increases polyphosphate levels that activate Lon protease degradation of the antitoxins, with subsequent release of active toxins (Maisonneuve et al., 2013). These free mRNAse toxins contribute to persistence to some, but not all antibiotics (Harms et al., 2016). The role of toxin action in bacteria of the phylum Firmicutes, and whether these toxins induce persistence or tolerance, is poorly understood. We therefore examined the role of *S. pyogenes* pSM19035-encoded  $\zeta$  toxin in growth arrest (dormancy), alone or with antibiotic in *B. subtilis* cells (representative of the Firmicutes), by controlling expression of the toxin at or near physiological concentrations. In our analysis, we did not study the role of (p)ppGpp in antitoxin degradation and free toxin release. We found that transient expression of a short-lived toxin  $\zeta$  variant ( $\zeta$ Y83C) induced different temporal sets of cell responses and growth arrest, but a small cell subpopulation ( $5 \times 10^{-5}$  to  $1 \times 10^{-4}$ ) exits the dormant state, leading to persistent or tolerant *B. subtilis* cells (Lioy et al., 2012).

Analysis of the metabolic changes induced by the free toxin showed that within the first 5 min,  $\zeta$ Y83C expression decreased the intracellular GTP pool and dysregulated transcription of 78 genes, of which 28 with reduced expression are essential for cell proliferation (Lioy et al., 2012). Induction of genes involved in the SOS response was not observed, but the expression was documented of genes that could modulate toxin action, such as increased *comGA* and *relA* expression or decreased *glmS* gene expression (Lioy et al., 2012). It is likely that by altering ATP:GTP ratios, toxin  $\zeta$ Y83C modifies availability of the initiating nucleotides; this in turn changes promoter preferences by RNA polymerase, and the intracellular signaling (Krasny and Gourse, 2004; Pedley and Benkovic, 2017).

Within the first 15 min of  $\zeta$ Y83C expression, the intracellular ATP concentration decreases and that of (p)ppGpp increases (Lioy et al., 2012). The contribution of increased *comGA* and *relA*

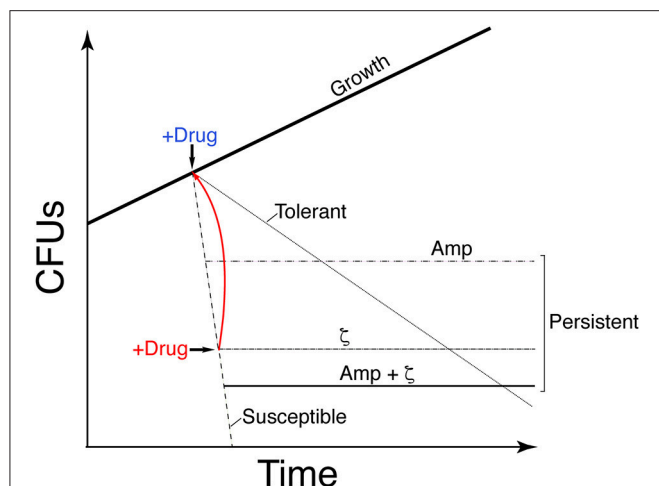
expression lead to higher (p)ppGpp levels (Potrykus and Cashel, 2008; Hahn et al., 2015; Liu et al., 2015), which directly inhibit both salvage and *de novo* GTP synthesis (Lopez et al., 1981; Kriel et al., 2012; Pedley and Benkovic, 2017). In *B. subtilis*, low GTP levels lead to derepression of CodY, a global transcriptional regulator from exponential to stationary phase (Handke et al., 2008; Kriel et al., 2012; Bittner et al., 2014; Brinsmade et al., 2014).

Downregulation of GlmS contributes indirectly to reducing the pool of UNAG synthesis, and a small UNAG pool increases levels of the essential cyclic 3,5-diadenosine monophosphate (c-di-AMP) second messenger (Witte et al., 2008; Zhu et al., 2016). Changes in the intracellular level of c-di-AMP, which play an essential role in K<sup>+</sup> transport and cell wall homeostasis (Gundlach et al., 2017), indirectly increase the intracellular (p)ppGpp pool (Rao et al., 2010; Corrigan et al., 2015). The relationship between the effective levels of c-di-AMP and bacterial persisters is nonetheless poorly characterized.

At later stages of toxin  $\zeta$ Y83C expression, synthesis of macromolecules (DNA, RNA, proteins) is inhibited and membrane potential is impaired (30–90 min; Lioy et al., 2012). Direct interaction of (p)ppGpp with DNA primase inhibits DNA replication (Wang et al., 2007; Srivatsan and Wang, 2008), (p)ppGpp-mediated low levels of GTP decrease mRNA transcription (Krasny and Gourse, 2004), and the essential GTPases decrease the amount of mature 70S ribosomes and reduce translation (Corrigan et al., 2016). Within 60–120 min, cell wall biosynthesis is reduced by  $\zeta$ -mediated phosphorylation of a UNAG fraction, leading to accumulation of unreactive UNAG-3P (Mutschler et al., 2011; Lioy et al., 2012), and by (p)ppGpp inhibition of peptidoglycan metabolism (Eymann et al., 2002). All these metabolic changes are reversible, however, because when the stress condition is relieved (or after artificial induction of antitoxin expression), the antitoxin  $\epsilon_2$  reverses the  $\zeta$ -induced dormant state and the cell population “awakens” (Tabone et al., 2014a,b).

When bacterial growth is challenged by addition of antibiotic, susceptible cells stop growing, but a small subpopulation shows persistence (a biphasic time-inactivation curve) or tolerance to the drug (a linear time-inactivation curve; see **Figure 1**; Lewis, 2010; Amato et al., 2013; Brauner et al., 2016). These complex phenotypes have been attributed to diverse stochastically induced stresses, with the toxin reducing the activity of the antibiotic or enhancing efflux activities to form persisters or tolerant cells (Lewis, 2010; Balaban et al., 2013; Brauner et al., 2016; Harms et al., 2016) or to produce cells susceptible to antibiotic action, as in *B. subtilis* (Wu et al., 2011; Tabone et al., 2014b). Toxin  $\zeta$  increases (p)ppGpp and decreases GTP pools, thus decreasing antibiotic persistence/tolerance formation; in contrast, low, dysregulated (p)ppGpp levels (in the  $\Delta$ *relA* context) increase toxin and antibiotic persistence/tolerance (Tabone et al., 2014b).

To analyze how toxin  $\zeta$  helps to induce a growth arrest state (dormancy), how antitoxin  $\epsilon_2$  promotes exit from this state, and to learn about the interconnection between toxin action and the persister/tolerant state, we have studied the metabolic activities of purine nucleotides on persister/tolerant bacterial. Transient controlled expression experiments with toxin and antitoxin showed that toxin  $\zeta$  induced a reversible growth-arrested state



**FIGURE 1** | Graphic illustration showing the difference in growth of the different stress survival strategies. Proliferation of susceptible clonal cells is halted by transient toxin  $\zeta$  expression (IPTG addition) or Amp addition (2x MIC) (+Drug, blue). A large fraction of cells is susceptible to the drug (dashed line); a subpopulation persists and forms colonies, leading to a biphasic time-inactivation curve ( $\zeta$  [dotdashed] or Amp [twodotted dashed] persisters) rather than a linear time-inactivation curve (tolerants; dotted line). Transient expression of antitoxin  $\varepsilon_2$  (+Drug, red) awakens the susceptible cells to toxin  $\zeta$  action (solid red line). Transient toxin  $\zeta$  expression and Amp addition yield distinct persister subpopulations.

in a large fraction of proliferating, susceptible *B. subtilis* cells, but that a small subpopulation persists rather than tolerating toxin action (see **Figure 1**). Controlled upregulation of antitoxin  $\varepsilon_2$  reversed growth arrest *in vivo* and inhibited the UNAG-dependent ATPase activity of toxin  $\zeta$  *in vitro*. GdpP- or DisA-dependent alteration of the c-di-AMP pool and CodY-dependent responses revealed that ampicillin (Amp) persisters and  $\zeta$ -mediated persisters are distinct subpopulations, perhaps with different exit control, and that Amp enhanced killing of  $\zeta$ -mediated persisters.

## MATERIALS AND METHODS

### Bacterial Strains and Plasmids

The bacterial strains and plasmids used in this study are listed in **Table 1**. All *B. subtilis* strains are isogenic with BG214. *Escherichia coli* BL21(DE3) cells harboring pBT290-borne  $\varepsilon$  gene under the transcriptional control of the T7 RNA polymerase-dependent promoter ( $P_{T7}$ ), or pCB920-borne wild type (*wt*)  $\zeta$  gene under the control of  $P_{T7}$  and  $\varepsilon$  gene under its native RNA polymerase  $\sigma^A$ -dependent promoter ( $P_\omega$ ) were used for protein purification as described (Camacho et al., 2002; Tabone et al., 2014b).

### Growth Conditions

The BG214 derivatives were grown to mid-exponential phase ( $\sim 5 \times 10^7$  cells  $\text{ml}^{-1}$ ) at  $37^\circ\text{C}$  in minimal medium S7 (MMS7) supplemented with the necessary amino acid (Lioy et al., 2006). Except for  $\Delta relA$ , strains were grown in MMS7 with methionine and tryptophan at  $50 \mu\text{g ml}^{-1}$  each (Lioy et al., 2006). The  $\Delta relA$

**TABLE 1** | Bacterial strains.

Strains	Relevant genotype	References
BG1125 <sup>a,b</sup>	+ <i>lacI</i> , $P_{hsp}$ $\zeta$ , <i>aadA</i> [pCB799- <i>xylR</i> , $P_{xylA}$ $\varepsilon$ , <i>cat</i> ]	Lioy et al., 2006
BG689 <sup>a</sup>	+ <i>xylR</i> , $P_{xylA}$ $\zeta$ Y83C, <i>cat</i>	Lioy et al., 2006
BG1145 <sup>a</sup>	+ $\Delta relA$ , <i>xylR</i> , $P_{xylA}$ $\zeta$ Y83C, <i>cat</i>	Lioy et al., 2012
BG1325 <sup>a</sup>	+ $\Delta gdpP$ , <i>xylR</i> , $P_{xylA}$ $\zeta$ Y83C, <i>cat</i>	This study
BG1323 <sup>a</sup>	+ $\Delta disA$ , <i>xylR</i> , $P_{xylA}$ $\zeta$ Y83C, <i>cat</i>	This study
BG1525 <sup>a</sup>	+ <i>codY::(erm::spc)</i> , <i>xylR</i> , $P_{xylA}$ $\zeta$ Y83C, <i>cat</i>	This study
BG1527 <sup>a</sup>	+ <i>codY::(erm::spc)</i> , $\Delta disA$ , <i>xylR</i> , $P_{xylA}$ $\zeta$ Y83C, <i>cat</i>	This study
BL21(DE3) <sup>c</sup>	+ [pCB920, $P_{T7}$ $\zeta$ gene, $P_\omega$ $\omega$ and $\varepsilon$ genes, <i>bla</i> ]	Tabone et al., 2014a
BL21(DE3) <sup>c</sup>	+ [pBT290, $P_{T7}$ $\varepsilon$ gene, <i>bla</i> ]	Camacho et al., 2002

<sup>a</sup>All *Bacillus subtilis* strains are isogenic with BG214 (*trpCE metA5 amyE1 ytsJ1 rsbV37 xre1 xkdA1 att<sup>SPB</sup> att<sup>CEB</sup>*).

<sup>b</sup>BG1125 cells bearing pCB799-borne  $\varepsilon$  gene were grown in MMS7 medium containing 0.05% xylose to titrate basal expression of the *wt*  $\zeta$  toxin.

<sup>c</sup>*Escherichia coli* BL21(DE3) genotype (*ompT gal* [ $\lambda$  DE3, *int::lacI::Plac<sub>UV5</sub>::T7 gene 1*] *fhuA2 [dcm]  $\Delta$ hdsS*).

strain shows an “auxotrophy phenotype” for valine, leucine, isoleucine and threonine, and was also supplemented with these amino acids ( $25 \mu\text{g ml}^{-1}$  each) (Roche, Germany; Lioy et al., 2006).

BG1125 bearing *lacI*- $P_{hsp}$  *wt*  $\zeta$  and pCB799-borne *xylR*- $P_{xylA}$  *wt*  $\varepsilon$  (**Table 1**), in which  $\zeta$  gene expression (transcribed by  $P_{hsp}$ ) is regulated by IPTG (Calbiochem, Spain) addition and the  $\varepsilon$  gene (transcribed by  $P_{xylA}$ ) is regulated by xylose (Xyl, Sigma, USA) addition (Lioy et al., 2012), was grown in MMS7 supplemented with Xyl (0.05%). In the absence of IPTG [Sigma, USA] there are  $\sim 40$   $\zeta$  toxin monomers/colony-forming units (CFU), which lead to genetic rearrangement. To titrate basal  $\zeta$  toxin levels, traces of Xyl (0.05%) were added to allow synthesis of low but marked  $\varepsilon_2$  antitoxin levels by the pCB799-borne  $\varepsilon$  gene. After IPTG addition, toxin  $\zeta$  concentration increased in a very short time (10 min) up to  $\sim 1,500$   $\zeta$  monomers/CFU, and its steady-state level remained for at least 240 min; these toxin levels are considered the “physiological concentration” (Lioy et al., 2012). At indicated times, 0.5% Xyl was added to induce antitoxin  $\varepsilon_2$  expression, and the culture was incubated 15 min before being plated without inducer or with 0.5% Xyl (Lioy et al., 2012).

In BG689 or BG1145 bearing the *xylR*- $P_{xylA}$   $\zeta$ Y83C cassette (**Table 1**), expression of the toxin  $\zeta$ Y83C variant was induced by addition of 0.5% Xyl. BG689 or BG1145 cells were grown in MMS7 to  $\sim 5 \times 10^7$  cells  $\text{ml}^{-1}$  at  $37^\circ\text{C}$ . Xylose addition increased  $\zeta$ Y83C levels to a plateau within the first 10 min, and the steady-state level of the toxin remained for at least 240 min (Tabone et al., 2014b).

Where indicated, toxin and/or antitoxin expression was induced by adding IPTG and/or Xyl. Before plating, cells were centrifuged and resuspended in fresh LB medium to remove the inducer or the antibiotic, and dilutions were plated on LB agar plates containing glucose (which switches off *xylR*- $P_{xylA}$  cassette expression) or Xyl to express the  $\varepsilon_2$  antitoxin. The

survival rate was derived from the number of CFU in a given condition relative to CFU of the non-induced/non-antibiotic-treated control. Except  $\Delta relA$ , cells grew in MMS7 with a doubling time of 50–60 min. The doubling time of  $\Delta relA$  cells increased 1.4-fold compared to the BG689 strain. Normal-sized and small colonies were observed in the  $\Delta relA$  and  $\Delta disA codY$  contexts. All plates were incubated for 20 h at 37°C.

The minimum inhibitory concentration (MIC) of Amp [Sigma, USA] was estimated by exposing  $1-3 \times 10^6$  cells  $\text{ml}^{-1}$  (16 h, 37°C) in MMS7 with shaking (240 rpm). The Amp concentration used ( $3 \mu\text{g ml}^{-1}$ ) was twice the MIC (2x MIC). In the absence of inducer, the presence of the  $\zeta Y83C$  (BG689 strain) or the  $\zeta$  gene (BG1125 bearing pCB799) does not affect the MIC (Tabone et al., 2014b).

## Protein Purification and Biochemical Assays

The *S. pyogenes* pSM19035-encoded  $\zeta$  gene was overexpressed in *E. coli* BL21(DE3) cells from a rifampicin-resistant T7 RNAP-dependent promoter as reported (Tabone et al., 2014a). In short, IPTG was added to induce the expression of T7 RNAP that transcribed *wt*  $\zeta$  toxin, and 30 min later rifampicin (Fluka, USA), was added to selectively block the expression of the  $\omega$  and  $\epsilon$  genes. After 120 min of incubation and full decay of the  $\epsilon_2$  antitoxin, the cells were harvested. The over-expressed long-living  $\zeta$  toxin was purified in two steps as described (Tabone et al., 2014a). The fractions containing the  $\zeta$  protein were dialyzed against buffer A (50 mM Tris-HCl pH 7.5, 80 mM NaCl) containing 50% glycerol and stored at  $-20^\circ\text{C}$ . The  $\epsilon$  gene was overexpressed in *E. coli* BL21(DE3) cells harboring pBT290 under the control of rifampicin-resistant  $P_{T7}$  (Ceglowski et al., 1993), and antitoxin  $\epsilon_2$  was overexpressed, and purified as described (Camacho et al., 2002). The purified protein was stored in buffer A containing 50% glycerol at  $-20^\circ\text{C}$  (Camacho et al., 2002).

The ATPase, dATPase or GTPase activities of  $\zeta$  toxin were measured using a (d)NTP/NADH-linked assay (De La Cruz et al., 2000; Yadav et al., 2012). Reactions (50  $\mu\text{l}$ ) contained the indicated concentration of  $\zeta$  toxin and the NADH enzyme mix (310  $\mu\text{M}$  NADH [Roche, Germany], 100 U  $\text{ml}^{-1}$  lactic dehydrogenase [Sigma, USA], 500 U  $\text{ml}^{-1}$  pyruvate kinase [Roche, Germany], and 2.5 mM phosphoenolpyruvate [Roche, Germany]) in buffer B (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MgOAc, 1 mM DTT, 50  $\mu\text{g/ml}$  BSA) with the indicated concentration of ATP, GTP or dATP, and 10 mM UNAG or uridine diphosphate-N-acetylgalactosamine (UNAGal) [Sigma, USA]. We determined the specific (d)NTPase activity (in  $\mu\text{M}$ ) by measuring the (d)NDP production rate using a Shimadzu CPS-20A dual-beam spectrophotometer as described (Yadav et al., 2012). A standard curve with known amounts of NADH was obtained and used to convert the rate of ADP/GDP/dADP production from absorbance/time to concentration/rate (De La Cruz et al., 2000; Yadav et al., 2012).

## RESULTS

### Toxin $\zeta$ Preferentially Hydrolyzes ATP

Toxin  $\zeta$  hydrolyzes ATP, even in the presence of a 10- to 15-fold excess of cold GTP (Tabone et al., 2014a), suggesting that

toxin  $\zeta$  prefers ATP to GTP (Tabone et al., 2014a). To examine these reactions, we purified toxin  $\zeta$  in the absence of its cognate antitoxin  $\epsilon_2$ .

In the absence of UNAG, toxin  $\zeta$  does not undergo autophosphorylation or hydrolyze NTP; with UNAG (2 mM) and 500 nM toxin  $\zeta$ , only traces of the  $\gamma$ -phosphate of ATP (0.5 mM) were transferred to UNAG (Tabone et al., 2014a). We tested directly for nucleotide used preferentially by toxin  $\zeta$ . Limiting  $\zeta$  concentrations (60 nM) were used to analyze  $\zeta$ -mediated ATP, GTP or dATP hydrolysis in physiological concentrations of UNAG and of nucleotides. The *B. subtilis* intracellular UNAG, ATP, GTP, and dATP pools approached  $\sim 10$ ,  $\sim 10$ ,  $\sim 5$  and  $\sim 0.02$  mM, respectively (Lopez et al., 1979; Lioy et al., 2012; Bittner et al., 2014).

Toxin  $\zeta$  did not hydrolyze purine nucleotide when UNAG was omitted (Figure 2A). At physiological UNAG and ATP concentrations (10 mM each), toxin  $\zeta$  (60 nM) hydrolyzed ATP in a reaction that rapidly reached saturation, which suggested that  $\zeta$  is a UNAG-dependent NTPase. The final rate of  $\zeta$  ATP hydrolysis approached the maximum rate ( $K_{\text{cat}}$ ) of  $1520 \pm 120 \text{ min}^{-1}$  (Figure 2A).

The UNAG-dependent  $\zeta$  ATPase activity was then compared with a *bona fide* ATPase enzyme. When the single-stranded DNA-dependent RecA ATPase was measured in parallel, *B. subtilis* RecA hydrolyzed ATP at near the previously observed  $K_{\text{cat}}$  of  $9 \pm 0.3 \text{ min}^{-1}$  (Yadav et al., 2014; Carrasco et al., 2015), which suggested that  $\zeta$  is a very robust ATPase. UNAG-dependent  $\zeta$ -mediated ATP hydrolysis was nonetheless sensitive to variations in ATP concentration, because when ATP was reduced to half (5 mM), the  $K_{\text{cat}}$  was reduced  $\sim 3$ -fold ( $510 \pm 44 \text{ min}^{-1}$ ).

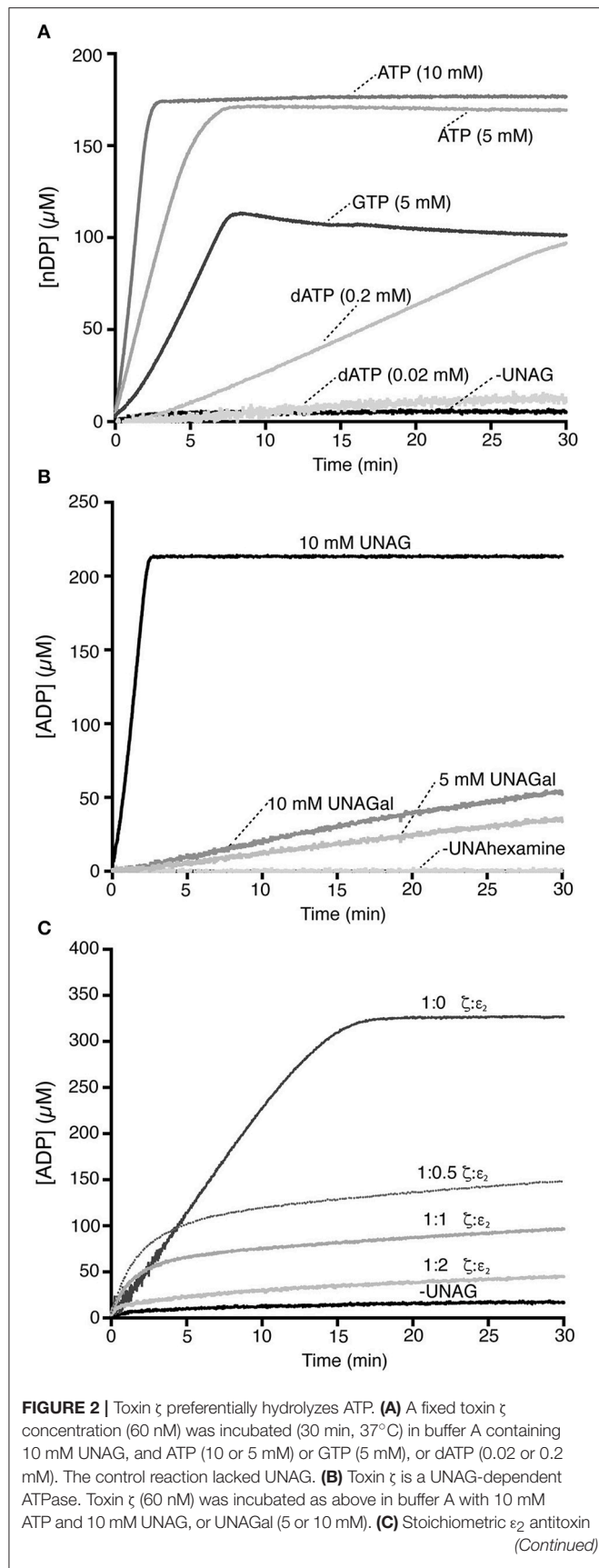
When ATP was replaced by physiological GTP concentrations (5 mM),  $\zeta$  was able to hydrolyze GTP in a UNAG-dependent manner and the reaction reached saturation in  $\sim 7$  min. The final steady state rate of GTP hydrolysis was reduced by  $\sim 5$ -fold ( $K_{\text{cat}} 280 \pm 47 \text{ min}^{-1}$ ) compared with physiological ATP concentrations (Figure 2A). Increasing the GTP concentration to 10 mM did not improve the reaction.

We analyzed the potential role of dATP as a substrate (Figure 2A). In the presence of physiological UNAG and dATP concentrations (10 and 0.02 mM, respectively), we observed no  $\zeta$ -mediated UNAG-dependent dATP hydrolysis (Figure 2A). To test whether  $\zeta$  catalyzes dATP hydrolysis, we increased its concentration artificially. At a 10-fold excess of dATP (0.2 mM),  $\sim 3$  min lag time was needed to reach the steady state rate of  $\zeta$ -mediated dATP hydrolysis; however, saturation was not reached in 30 min reaction (Figure 2A). With a 10-fold excess of dATP, its hydrolysis was reduced by  $\sim 20$ -fold compared with ATP. ATP is probably the preferred  $\zeta$  nucleotide cofactor.

### UNAGal Is a Poor Inducer of the Toxin $\zeta$ ATPase

Toxin  $\zeta$  interacts specifically with UNAG rather than UDP-glucose (Mutschler et al., 2011); in addition, *B. subtilis* GalE is able to interconvert UNAG and UDP-N-acetylgalactosamine (UNAGal), and the cell wall contains N-acetylglucosamine and N-acetylgalactosamine (Soldo et al., 2003). To determine whether UNAGal, a C-4 epimer of UNAG, can activate toxin  $\zeta$  ATPase



**FIGURE 2 |** Continued

inhibits the UNAG-dependent  $\zeta$  ATPase. A fixed  $\zeta$  toxin concentration (30 nM) and increasing antitoxin  $\epsilon_2$  concentrations (15–60 nM) were incubated (30 min, 37°C) in buffer A containing limiting concentrations of ATP (2 mM) and UNAG (4 mM). The amount of ATP hydrolyzed was calculated (see Section Materials and Methods). The control reaction lacks UNAG. All reactions were repeated three or more times with similar results.

activity, we carried out ATPase assays with increasing UNAGal concentrations.

In the absence of UNAG or UNAGal (minus UNAhexamines), ATP hydrolysis by toxin  $\zeta$  was at background level (**Figure 2B**). Quantitative analysis of these reactions showed that at physiological UNAGal concentrations, the final  $\zeta$ -mediated ATP hydrolysis rate was  $\sim 85$ -fold lower ( $K_{cat}$  20 min<sup>-1</sup>) than  $\zeta$  in the presence of UNAG. In the presence of a UNAGal excess (10 mM), the  $K_{cat}$  was slightly increased (28 min<sup>-1</sup>), but was still  $\sim 60$ -fold lower than that at physiological UNAG concentrations (**Figure 2B**); this result indicates that  $\zeta$  ATPase activity is specifically stimulated by UNAG rather than by UNAGal. PetZ similarly accumulates UNAG-3P after 60 min, and UNAGal-3P after 720 min incubation (Mutschler et al., 2011).

### Antitoxin $\epsilon_2$ Inhibits UNAG-Dependent $\zeta$ -Mediated ATP Hydrolysis

*In vitro*, the  $\zeta\epsilon_2\zeta$  complex is reported to hydrolyze ATP and phosphorylate UNAG to form inactive UNAG-3P (Mutschler et al., 2011). In contrast, *in vivo* experiments showed that the  $\epsilon_2$  antitoxin inhibits the effect of toxin  $\zeta$ , perhaps by forming the inactive  $\zeta\epsilon_2\zeta$  complex (Lioy et al., 2006, 2010). To test whether toxin  $\zeta$  hydrolyzes ATP in the presence of the antitoxin  $\epsilon_2$ , both proteins were purified separately (Camacho et al., 2002; Tabone et al., 2014a) and UNAG-dependent ATPase activity measured.

The antitoxin  $\epsilon_2$ , alone or with UNAG, did not hydrolyze ATP (**Figure 2C**). In the presence of UNAG and ATP, the rate of UNAG-dependent  $\zeta$ -mediated ATP hydrolysis was reduced by increasing antitoxin  $\epsilon_2$  concentrations (**Figure 2C**). At  $\zeta:\epsilon_2$  ratios of 1:0.5 or 1:1, the kinetics of  $\zeta$ -mediated ATP hydrolysis was initially unaltered, but ATP hydrolysis was inhibited after 5 min. At a slight  $\epsilon_2$  excess (1:2 ratio), the antitoxin inhibited  $\zeta$  ATPase activity (**Figure 2C**). Results were similar when both proteins were preincubated (5 min) at a 1:1  $\zeta:\epsilon_2$  ratio ( $\zeta\epsilon_2\zeta$  complex; not shown), which suggests that when it interacts with  $\zeta$ , the antitoxin occupies the ATP binding pocket (Meinhart et al., 2003) and inhibits toxin ATPase activity. This is consistent with the crystal structure of the biologically inactive  $\zeta\epsilon_2\zeta$  complex and with the interpretation that antitoxin  $\epsilon_2$  is necessary and sufficient to inactivate toxin  $\zeta$ . It is likely that the long reaction incubation time (24 h) and/or low  $\epsilon_2$  stability could explain discrepancies with the previous report (Mutschler et al., 2011).

### Toxin $\zeta$ Induces Reversible Growth Arrest But a Small Subpopulation Evades Its Action

The release of toxins from their cognate antitoxins [or induction of toxin expression (+Drug in blue in **Figure 1**)], should lead



to a bimodal time-inactivation curve if persisters appeared (dotdashed line). This deviates from the simple decay, anticipated for a population of only susceptible cells (dashed line) or for a uniformly tolerant bacterial population (dotted line, in **Figure 1**; Brauner et al., 2016; Harms et al., 2016). Inactivation of the toxin by expression of the antitoxin (+Drug in red) should lead to recovery of the plating efficiency (red solid line) if the toxin is bacteriostatic (**Figure 1**). To test whether expression of physiological levels of free toxin  $\zeta$  induce persistence (dotdashed line) or tolerance (dotted line) and to study the mechanism used for such a phenotype (bacteriostasis or bacteriolysis) we performed long term survival assays. Toxin  $\zeta$  was induced for a long period, and then antitoxin  $\varepsilon_2$  expression was induced (**Figure 3A**).

*Bacillus subtilis* BG1125 bearing the  $\zeta$  gene under the control of IPTG induction is prone to rearrangement in the absence of IPTG (Lioy et al., 2012). To overcome this effect, the pCB799-borne  $\varepsilon$  gene under the control of Xyl was transferred into the background (**Table 1**; see Section Materials and Methods). BG1125 cells bearing pCB799 were grown in MMS7 supplemented with 0.05% Xyl, to  $\sim 5 \times 10^7$  cells  $\text{ml}^{-1}$  ( $\text{OD}_{560} = 0.2$ ), and expression of the  $\zeta$  gene was induced by IPTG addition (time zero). Cells, which formed colonies after plating on LB agar without IPTG, showed a bimodal time-inactivation curve suggesting the presence of persisters (**Figure 3A**), rather than showing a uniform simple decay, expected for tolerant cells (**Figure 1**, dotted line).

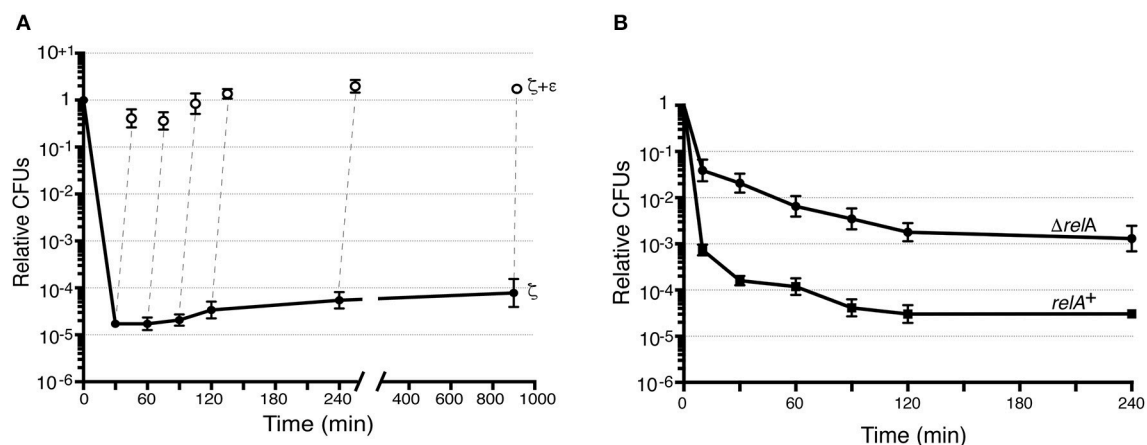
To test whether the persisters are due to noise that causes instability in a bacterial population (a reduced cell fraction transiently insensitive to toxin action) or noisy gene expression (a reduced fraction with no toxin expression), we maintained IPTG induction up to 900 min, after which cells were plated in the absence of the inducer. In the former case, only a

fraction of the non-replicating dormant cells would exit the arrest state and resume growth after plating without IPTG, whereas in the latter case, cell proliferation of persisters is predicted to increase 8- to 16-fold. After IPTG addition, the small persister subpopulation increased slightly ( $\sim 3$ -fold) during the first 240 min, to later remain apparently constant (**Figure 3A**); this suggested negligible biological noise during the first 240 min, and persisters were transiently insensitive to toxin action.

Massive expression of toxin PezT or  $\zeta$  triggers an irreversible bactericidal effect in *E. coli* grown in rich medium or *B. subtilis* grown in minimal medium, respectively (Zielenkiewicz and Ceglowski, 2005; Mutschler et al., 2011), but physiological concentrations of free toxin  $\zeta$  induce a reversible bacteriostatic state (Lioy et al., 2012). To identify the source of these discrepancies, we tested whether IPTG-induced growth arrest in *B. subtilis* cells is fully reversible after antitoxin  $\varepsilon_2$  expression triggered by 0.5% Xyl (15 min), followed by plating in LB agar with 0.5% Xyl but lacking IPTG. Antitoxin  $\varepsilon_2$  expression was sufficient to reverse growth arrest, and most cells recovered proliferation, even after 900 min of toxin  $\zeta$  action (**Figure 3A**). Although, a reduced fraction (10 to 15% of total cells) were stained with propidium iodide, suggesting a membrane compromise in these cells. It is likely that toxin  $\zeta$  induces a reversible inhibition of cell growth, and that antitoxin  $\varepsilon_2$  expression is necessary and sufficient to switch off toxin-induced responses, with cells awakening and forming colonies even after 15 h of toxin incubation (see **Figure 1**, solid red line), but 10 to 15% of total cells might loss cell viability.

## Dysregulated (p)ppGpp Levels Increase the Rate of $\zeta$ Y83C Persisters

*Bacillus subtilis* encodes one bifunctional RelA synthase-hydrolase and two mono-functional SasA (also termed



**FIGURE 3 |** Toxin  $\zeta$  induces reversible dormancy and selects for pre-existing persisters. **(A)** BG1125 cells (*lacI-P<sub>hsp</sub>  $\zeta$  spc* cassette) bearing pCB799-borne *xylR-P<sub>XylA</sub>  $\varepsilon$*  cassette were grown in MMS7 medium containing traces of xylose (Xyl; 0.05%) to  $\sim 5 \times 10^7$  cells  $\text{ml}^{-1}$  (37°C). IPTG (2 mM) was added to half the culture to induce  $\zeta$  expression (time 0) and the culture was further incubated. At various times, samples were withdrawn and plated in LB agar plates (●,  $\zeta$ -expressing) or to allow antitoxin expression, 0.5% Xyl was added, the culture incubated (15 min) and plated onto LB agar plates (○, antitoxin  $\varepsilon_2$  induction). **(B)** The effect of  $\zeta$ Y83C expression on CFU was measured. BG689 (■) or BG1145 (●) cells were cultured in MMS7 to  $\sim 5 \times 10^7$  cells  $\text{ml}^{-1}$  (37°C). Xyl (0.5%) was added to half of the culture to induce  $\zeta$ Y83C expression (time 0). At various times, samples were withdrawn and plated onto LB agar plates. Data are shown as mean  $\pm$  standard error of the mean (SEM), from  $>4$  independent experiments.

YwaC/RelP/Sas1) and SasB (YjbM/RelQ/Sas2) synthases [see Nanamiya et al., 2008; Srivatsan and Wang, 2008]. In the absence of RelA, an excess of GTP over GDP as well as baseline levels of (p)ppGpp “dysregulated” by the SasA and SasB synthases, increase toxin persistence or tolerance by >150-fold (Tabone et al., 2014b). This effect correlates with dysregulated (p)ppGpp levels. Lowering the GTP levels without affecting (p)ppGpp, by treating cells with decoyinine (a GMP synthetase inhibitor), the persistent rate was indistinguishable between treated or untreated  $\Delta relA$  cells (Lioy et al., 2012).

To test whether the CFU increase correlates with a simple decay curve predicted from a uniform tolerant bacterial population or with a biphasic time-inactivation curve due to persistence (Figure 1), we analyzed toxin expression in the  $relA^+$  (BG689) or  $\Delta relA$  (BG1145) cells bearing the  $xylR-P_{xylA}$   $\zeta$ Y83C cassette (Table 1). The  $relA^+$  and  $\Delta relA$  cells were grown in MMS7 to  $\sim 5 \times 10^7$  ml<sup>-1</sup>, expression of the  $\zeta$ Y83C gene was induced with 0.5% Xyl, and the time-inactivation curve was analyzed. In the absence of RelA, a typical biphasic curve was observed upon expression of physiological concentrations of the toxin, with an  $\sim 160$ -fold ( $\sim 5 \times 10^{-3}$ ) increase in the rate of persisters after plating on LB agar without Xyl (Figure 3B).

## Varying the c-di-AMP Pool Alters the Rate of Toxin But Not of Amp Persistence

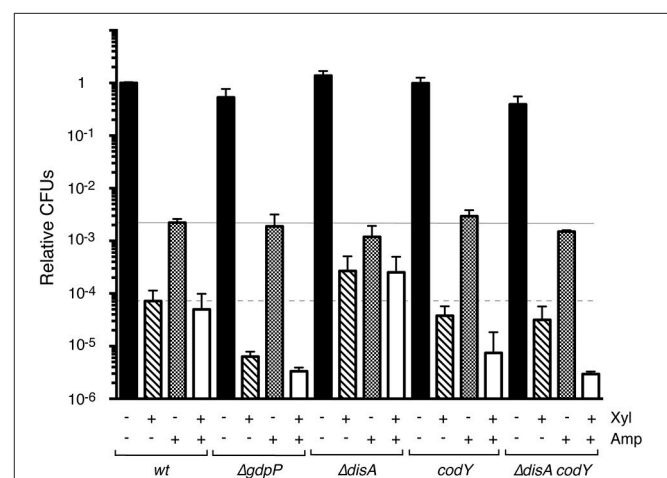
The second messenger c-di-AMP, which is at the heart of cell wall homeostasis, is produced mainly by Gram-positive bacteria of the phyla Firmicutes and Actinobacteria, and by some species of the  $\delta$ -Proteobacteria class (Corrigan and Grundling, 2013). In Firmicutes, high or low c-di-AMP levels indirectly increase (p)ppGpp (Rao et al., 2010; Corrigan et al., 2015), whereas in *Staphylococcus aureus*, they, respectively, increase or decrease  $\beta$ -lactam tolerance/resistance (Corrigan et al., 2011, 2015). It is not known whether c-di-AMP has a role in toxin  $\zeta$  responses to stress.

In Firmicutes, intracellular c-di-AMP levels are precisely regulated by two sets of enzymes with opposite effects and by two purine nucleotides. The diadenylate cyclases (DAC) synthesize c-di-AMP from two ATP molecules, the phosphodiesterases (PDE) degrade c-di-AMP into pApA; (p)ppGpp and pApA inhibit PDE enzyme activity (Rao et al., 2010; Corrigan and Grundling, 2013; Huynh and Woodward, 2016), which predicts that c-di-AMP levels increase during starvation. Exponentially growing *B. subtilis* cells express two DAC (DisA and CdaA) and two PDE enzymes (GdpP and PgpH; Rao et al., 2010; Corrigan et al., 2011; Corrigan and Grundling, 2013; Commichau et al., 2015; Huynh and Woodward, 2016). The absence of both DAC or of both PDE causes aberrant physiology and synthetic lethality when the medium contained high K<sup>+</sup> (5 mM KCl), but one representative of each family can be deleted with no apparent effect (Corrigan et al., 2011; Corrigan and Grundling, 2013; Commichau et al., 2015; Huynh and Woodward, 2016; Gundlach et al., 2017). C-di-AMP levels vary marginally (2- to 3-fold) in cells lacking DisA, CdaA or GdpP compared to the *wt* strain (Oppenheimer-Shaanan et al., 2011; Gándara and Alonso, 2015).

To determine how *B. subtilis* cells respond to toxin-mediated stress, we induced transient toxin  $\zeta$ Y83C expression and studied

the effect of disturbing the c-di-AMP metabolic balance by deleting one DAC (DisA) or one PDE (GdpP) enzyme on an isogenic background (Table 1). In parallel, Amp was used as a second stressor at twice the MIC, in anticipation that toxin expression and Amp would respond to different physiological cues. The MIC of Amp was similar in all strains tested. After Amp exposure, a subpopulation of clonal cells yielded a biphasic time-kill curve (twodotted dashed line), which indicated that they persisted rather than becoming Amp-tolerant (dotted line; Figure 1). Similar biphasic curves are reported for other bacterial species after Amp exposure (Lewis, 2010; Amato et al., 2013; Brauner et al., 2016; Harms et al., 2016).

*Bacillus subtilis* cells that lack GdpP show intracellular c-di-AMP levels  $\sim 2$ -fold higher than those of the *wt* strain (Gándara and Alonso, 2015). Absence of GdpP indirectly increases (p)ppGpp pools (Gundlach et al., 2015; Zhu et al., 2016), which suggests that a small number of specific signaling nucleotides integrate and coordinate key metabolic intersections in response to variation of the intracellular c-di-AMP pool. We constructed and analyzed a strain bearing the  $xylR-P_{xylA}$   $\zeta$ Y83C cassette in the context of  $\Delta gdpP$  (Table 1). Regulated  $\zeta$ Y83C expression in the *wt* or  $\Delta gdpP$  contexts produced a typical biphasic survival curve, with an initial rapid decrease in CFU and a persistent subpopulation with a stable number of CFU (between 10 and 300 min; not shown); for direct comparison of the various strains, data at 120 min are shown (Figure 4). In the *wt* strain,  $\zeta$ Y83C expression used (p)ppGpp to mediate rapid inhibition of cell proliferation, and a cell subpopulation entered a toxin- ( $\sim 7.2 \times 10^{-5}$ ) or Amp-persistent state ( $\sim 2.1 \times 10^{-3}$ ; Figure 4), as reported (Lioy et al., 2012). In the  $\Delta gdpP$  strain, after transient toxin expression, the persistence rate decreased by  $\sim 10$ -fold ( $7 \times 10^{-6}$ ), but did not significantly affect the persistence rate after



**FIGURE 4 |** Variation in the c-di-AMP or GTP pools alters the level of  $\zeta$ Y83C toxin persists, but not stochastic Amp persists during exponential growth. The set of isogenic strains was grown in MMS7 to  $\sim 5 \times 10^7$  cells ml<sup>-1</sup> (37°C), followed by addition of 0.5% Xyl and/or Amp (3  $\mu$ g ml<sup>-1</sup>). Cultures were incubated (120 min), then plated onto LB agar plates. The number of CFU is shown relative to the non-induced/non-Amp-treated control. + and - denote the presence or absence of the indicated compound. Data shown as mean  $\pm$  SEM, from >4 independent experiments.

Amp addition ( $\sim 1.7 \times 10^{-3}$ ). In the absence of GdpP, exposure to Amp and Xyl decreased the rate of persisters by  $\sim 16$ -fold ( $\sim 3 \times 10^{-6}$  survivals) compared to the *wt* strain (**Figure 4**), which suggests that a subset of toxin or Amp persisters randomly switched to the susceptible state and were targeted by Amp or the toxin.

Cells lacking DisA have  $\sim 2$ -fold lower levels of intracellular c-di-AMP than *wt* cells (Oppenheimer-Shaanan et al., 2011; Gándara and Alonso, 2015). We constructed the *xylR-P<sub>xylA</sub> ζY83C ΔdisA* strain (**Table 1**), and found that the persister cell rate was slightly affected by Amp addition ( $< 2$ -fold,  $\sim 1.5 \times 10^{-3}$  survivals). Transient toxin ζY83C expression increased persisters  $\sim 4$ -fold ( $\sim 3 \times 10^{-4}$ ) in the *ΔdisA* compared to the *wt* strain (**Figure 4**). Transient toxin ζY83C expression and Amp addition did not notably affect colony formation compared to addition of Xyl alone (**Figure 4**).

### Absence of CodY Alters the Rate of Toxin ζY83C But Not of Amp Persistence

Transient toxin ζ expression decreased the GTP pool in exponentially growing *B. subtilis* cells (Liroy et al., 2012). Intracellular GTP levels have a central role in modulating the stringent response and in reprogramming gene regulation to allow appropriate adaptation to stress. CodY, a GTP-binding protein, is a pleiotropic regulator that senses intracellular branched chain amino acids and GTP levels (Sonenshein, 2007). Low GTP levels, as found during acute stress, release CodY from DNA, leading to deregulation of genes involved in adaptation to nutrient limitation (Ratnayake-Lecamwasam et al., 2001; Belitsky and Sonenshein, 2013; Bittner et al., 2014). The role of CodY in toxin ζ stress responses is unknown. To test whether CodY modulates toxin and/or antibiotic persistence, we constructed the *xylR-P<sub>xylA</sub> ζY83C codY* strain (**Table 1**). Lack of CodY did not markedly alter the Amp persister rate ( $\sim 2.5 \times 10^{-3}$  survivals; **Figure 4**). After Xyl exposure, however, we observed a slight decrease in the toxin persister rate ( $\sim 2$ -fold) compared to *wt* cells (**Figure 4**). Transient Xyl and Amp addition decreased CFU ( $\sim 8 \times 10^{-6}$  survivals) and the level of persisters decreased by  $\sim 5$ -fold compared to the *wt* strain (**Figure 4**), which suggest that cells lacking CodY adapt poorly to toxin and Amp stress.

### Lack of CodY Suppresses Toxin Persistence Triggered by Low c-di-AMP Levels

The absence of DisA increased, and of CodY or GdpP decreased the rate of toxin persistence (**Figure 4**). Since lack of both CodY and GdpP strongly affected cell recovery, but the combined absence of CodY and DisA showed a less stringent phenotype, we constructed the *xylR-P<sub>xylA</sub> ζY83C ΔdisA codY* strain (BG1527; **Table 1**). The BG1527 strain yielded colonies with diffuse borders, a 3:1 normal:small size ratio, and viability reduced by  $\sim 1.4$ -fold compared to parental BG689 strains, but lack of CodY and DisA did not notably alter the rate of Amp persisters ( $\sim 2.5 \times 10^{-3}$  survivals). Following toxin ζY83C expression, we observed a moderate decrease ( $\sim 2$ -fold) in the toxin persister rate ( $\sim 3.5 \times 10^{-5}$  survivals) compared to *wt* cells, similar to the *codY* strain (**Figure 4**). Addition of Xyl and Amp greatly decreased

the persistence rate ( $\sim 3 \times 10^{-6}$  survivals; **Figure 4**). Different clonal subpopulations of persisting cells thus probably evolved differently in response to the toxin and Amp.

## DISCUSSION

Toxin ζ represents a class of UNAG-dependent ATPases (**Figure 2A**). As another mechanism to halt cell proliferation, toxin ζ also catalyzes the transfer of part of the ATP γ-phosphate generated upon ATP hydrolysis to a fraction of UNAG, to yield unreactive UNAG-3P (Mutschler et al., 2011; Tabone et al., 2014a). Stoichiometric concentrations of purified antitoxin ε<sub>2</sub> are necessary and sufficient to inactivate toxin ζ action, which suggests that no other factor contributes to ζ inactivation *in vitro*.

Using a set of isogenic *B. subtilis* strains, we tested how purine nucleotide signaling integrates and coordinates the toxin mode of action *in vivo*. Toxin ζ expression induced a biphasic time-inactivation curve with initial rapid, reversible growth arrest of the bulk of susceptible cells; a minor cell subpopulation showed non-inheritable toxin persistence rather than tolerance. Subsequent expression of the ε<sub>2</sub> antitoxin reversed ζ-induced dormancy, and the cells formed colonies even after 900 min of growth arrest. After accumulation of the ζε<sub>2</sub> complex, the heterogeneous dormancy state is nearly fully reversible; but a reduced subpopulation (up to 15%) of total cells is still stained with propidium iodide. It is likely that the ζ phosphotransferase might compromise the awakening of these cells, which may have a poor fitness or be maladapted. Persisters are formed through redundant mechanisms, and both the biological basis of persistence and the mechanisms that lead to persister formation are poorly understood in Firmicutes. Direct comparison with the well-characterized *E. coli* system could introduce some noise. For example, in both *E. coli* and *B. subtilis* cells, physiological (p)ppGpp levels are necessary for toxin-induced persistence (Korch et al., 2003; Nguyen et al., 2011; Liroy et al., 2012; Amato et al., 2013; Maisonneuve et al., 2013). In the absence of hydrolase-synthase SpoT, *E. coli* cells are not viable (Xiao et al., 1991), but in the *spoT1* context (attenuated hydrolase activity), high levels of dysregulated (p)ppGpp give rise to hypertolerance (Amato et al., 2013; Maisonneuve et al., 2013). In *B. subtilis* cells, lack of the hydrolase-synthase RelA leads to undetectable levels of dysregulated (p)ppGpp, which in turn do not inhibit GTP synthesis and contribute indirectly to hyperpersistence ( $\sim 160$ -fold increase). In the absence of (p)ppGpp, there is no persistence signal in *B. subtilis* or *S. aureus* cells, but reduction of GTP and ATP (Tabone et al., 2014b) or ATP levels (Conlon et al., 2016), respectively, leads to cell susceptibility to distinct antibiotics. Indeed, the artificial reduction of the GTP level sensitizes the cells to different antibiotics in the absence of (p)ppGpp (Tabone et al., 2014b). In *E. coli* cells that lack the 10 host-encoded mRNA interferases, levels of persisters to certain antibiotics decrease (Maisonneuve et al., 2011), whereas in *B. subtilis* cells, absence of the single mRNAase NdoA (MazF) increases antibiotic lethality rather than inducing persister cell formation (Wu et al., 2011).

A very small fraction of *E. coli* cells ( $\sim 0.01\%$ ) is reported to have a high (p)ppGpp concentration, which triggers entry into the persistent state (Maisonneuve et al., 2013). Our study



addressed the mechanism of persister formation in *B. subtilis* cells in conditions in which toxin/antitoxin expression were controlled by external inducers; antitoxin degradation thus had no role, which rendered unnecessary the analysis of (p)ppGpp in toxin release. If the stochastic switch to produce (p)ppGpp is the sole factor that triggers persister formation, the proportion of toxin and Amp persisters should be similar, and transient toxin expression and Amp addition would not further decrease cell viability. This was not observed. To explain our results, we must assume that a certain cell fraction switches stochastically to the persistent state prior to environmental challenges (Amp persisters), but sensing the metabolic state is of key importance for responsive induction of persistence. Alterations in the GTP (*codY*) or c-di-AMP pools (*gdpP* or *disA*) indicated a constant “awakening” rate after Amp addition, but a variable proportion of toxin persisters (Figure 4). Toxin  $\zeta$  temporarily and reversibly increases the (p)ppGpp pool, reduces the ATP and GTP pools, and modulates c-di-AMP and UNAG levels to allow cells to readjust their metabolism from logarithmic growth to “growth arrest,” enabling cells to cope with environmental stress. The pattern of toxin persistence was varied by altering the c-di-AMP pool, which acts as two opposite mechanisms that are negatively and positively controlled by toxin expression. The subpopulation of bet-hedging persister cells that arises before changes in the environment and those triggered by toxin-induced metabolic changes both lead to toxin persisters.

Responsive strategies based on environmental sensing alter phenotypic switching between growth-arrested and persister cells. By varying the intracellular pool of signaling nucleotide, the stochastic subpopulation of toxin  $\zeta$  persisters varied up to 40-fold ( $\Delta$ *disA* vs.  $\Delta$ *gdpP*). When both stress sources (Amp and free toxin) were present, however, a fraction of Amp (or toxin) persisters might awaken and become sensitive to the second stressor, decreasing the rate of persisters by up to 200-fold ( $\Delta$ *disA* vs.  $\Delta$ *gdpP* background).

Based on these results and our previous work (Lioy et al., 2006, 2012; Tabone et al., 2014a,b), we propose that the *modus operandi* of toxin  $\zeta$ -induced growth arrest is to reduce

the ATP (by direct hydrolysis) and GTP (by conversion to [p]ppGpp) pools. As a consequence of this, (p)ppGpp levels are increased, and a fraction of UNAG becomes phosphorylated. High (p)ppGpp directly inhibits regeneration and *de novo* GTP synthesis; it positively and negatively regulates the c-di-AMP pool, and decreases the proton-motive force (lowering the ATP pool) as well as UNAG (Kriel et al., 2012). These imbalances induce diverse transient, reversible states to ensure population survival in adverse conditions. Except (p)ppGpp dysregulation on the  $\Delta$ *relA* background, there is no direct information that a discrete metabolite increases persister formation. ATP depletion is thought to be a general mechanism of persister formation in bacteria (Conlon et al., 2016; Shan et al., 2017), although we found that a reduction in ATP levels leads to  $\zeta$ -induced growth arrest rather than to persister formation. We propose that the interrelationship between ATP, GTP, (p)ppGpp, c-di-AMP, and UNAG contribute, via a poorly characterized mechanism, to  $\zeta$ -induced cell growth arrest and persister formation.

## AUTHOR CONTRIBUTIONS

MM, VL, MT, and JA conceived and designed the experiments for this study. MM, VL, and MT performed the experiments. JA wrote the manuscript. All authors discussed the data and made comments on the manuscript.

## ACKNOWLEDGMENTS

We thank Boris R. Belitsky (Tufts University School of Medicine, Boston, MA USA) for providing the BB1043 (*codY::[erm::spe]*) mutant strain, Silvia Ayora (CNB-CSIC) for critical comments, and Catherine Mark (CNB-CSIC) for editorial assistance. MT is a PhD fellow of the La Caixa Foundation International Fellowship Programme (La Caixa/CNB). This study was supported by the Spanish Ministerio de Economía y Competitividad and the European Union (MINECO-FEDER) BFU2015-67065-P to JA.

## REFERENCES

- Amato, S. M., Orman, M. A., and Brynildsen, M. P. (2013). Metabolic control of persister formation in *Escherichia coli*. *Mol. Cell* 50, 475–487. doi: 10.1016/j.molcel.2013.04.002
- Balaban, N. Q., Gerdes, K., Lewis, K., and McKinney, J. D. (2013). A problem of persistence: still more questions than answers? *Nat. Rev. Microbiol.* 11, 587–591. doi: 10.1038/nrmicro3076
- Belitsky, B. R., and Sonenshein, A. L. (2013). Genome-wide identification of *Bacillus subtilis* CodY-binding sites at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7026–7031. doi: 10.1073/pnas.1300428110
- Bittner, A. N., Kriel, A., and Wang, J. D. (2014). Lowering GTP level increases survival of amino acid starvation but slows growth rate for *Bacillus subtilis* cells lacking (p)ppGpp. *J. Bacteriol.* 196, 2067–2076. doi: 10.1128/JB.01471-14
- Brauner, A., Fridman, O., Gefen, O., and Balaban, N. Q. (2016). Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* 14, 320–330. doi: 10.1038/nrmicro.2016.34
- Brinsmade, S. R., Alexander, E. L., Livny, J., Stettner, A. I., Segre, D., Rhee, K. Y., et al. (2014). Hierarchical expression of genes controlled by the *Bacillus subtilis* global regulatory protein CodY. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8227–8232. doi: 10.1073/pnas.1321308111
- Camacho, A. G., Misselwitz, R., Behlke, J., Ayora, S., Welfle, K., Meinhart, A., et al. (2002). *In vitro* and *in vivo* stability of the  $e_2\zeta_2$  protein complex of the broad host-range *Streptococcus pyogenes* pSM19035 addiction system. *Biol. Chem.* 383, 1701–1713. doi: 10.1515/bc.2002.191
- Carrasco, B., Yadav, T., Serrano, E., and Alonso, J. C. (2015). *Bacillus subtilis* RecO and SsbA are crucial for RecA-mediated recombinational DNA repair. *Nucleic Acids Res.* 43, 5984–5997. doi: 10.1093/nar/gkv545
- Ceglowski, P., Boitsov, A., Chai, S., and Alonso, J. C. (1993). Analysis of the stabilization system of pSM19035-derived plasmid pBT233 in *Bacillus subtilis*. *Gene* 136, 1–12. doi: 10.1016/0378-1119(93)90441-5
- Commichau, F. M., Dickmanns, A., Gundlach, J., Ficner, R., and Stulke, J. (2015). A jack of all trades: the multiple roles of the unique essential second messenger cyclic di-AMP. *Mol. Microbiol.* 97, 189–204. doi: 10.1111/mmi.13026
- Conlon, B. P., Rowe, S. E., Gandt, A. B., Nuxoll, A. S., Donegan, N. P., Zalis, E. A., et al. (2016). Persister formation in *Staphylococcus aureus* is associated with ATP depletion. *Nat. Microbiol.* 1, 16051. doi: 10.1038/nrmicrobiol.2016.51



- Corrigan, R. M., Abbott, J. C., Burhenne, H., Kaever, V., and Grundling, A. (2011). c-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. *PLoS Pathog.* 7:e1002217. doi: 10.1371/journal.ppat.1002217
- Corrigan, R. M., Bellows, L. E., Wood, A., and Grundling, A. (2016). ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1710–E1719. doi: 10.1073/pnas.1522179113
- Corrigan, R. M., Bowman, L., Willis, A. R., Kaever, V., and Grundling, A. (2015). Cross-talk between two nucleotide-signaling pathways in *Staphylococcus aureus*. *J. Biol. Chem.* 290, 5826–5839. doi: 10.1074/jbc.M114.598300
- Corrigan, R. M., and Grundling, A. (2013). Cyclic di-AMP: another second messenger enters the fray. *Nat. Rev. Microbiol.* 11, 513–524. doi: 10.1038/nrmicro3069
- De La Cruz, E. M., Sweeney, H. L., and Ostap, E. M. (2000). ADP inhibition of myosin V ATPase activity. *Biophys. J.* 79, 1524–1529. doi: 10.1016/S0006-3495(00)76403-4
- Eymann, C., Homuth, G., Scharf, C., and Hecker, M. (2002). *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *J. Bacteriol.* 184, 2500–2520. doi: 10.1128/JB.184.9.2500-2520.2002
- Gándara, C., and Alonso, J. C. (2015). DisA and c-di-AMP act at the intersection between DNA-damage response and stress homeostasis in exponentially growing *Bacillus subtilis* cells. *DNA Repair* 27, 1–8. doi: 10.1016/j.dnarep.2014.12.007
- Gerdes, K. (2013). *Prokaryotic Toxin-antitoxins*. Berlin: Springer. doi: 10.1007/978-3-642-33253-1
- Gundlach, J., Herzberg, C., Kaever, V., Gunka, K., Hoffmann, T., Weiss, M., et al. (2017). Control of potassium homeostasis is an essential function of the second messenger cyclic di-AMP in *Bacillus subtilis*. *Sci. Signal.* 10:eal3011. doi: 10.1126/scisignal.aal3011
- Gundlach, J., Mehne, F. M., Herzberg, C., Kampf, J., Valerius, O., Kaever, V., et al. (2015). An essential poison: synthesis and degradation of cyclic Di-AMP in *Bacillus subtilis*. *J. Bacteriol.* 197, 3265–3274. doi: 10.1128/JB.00564-15
- Hahn, J., Tanner, A. W., Carabetta, V. J., Cristea, I. M., and Dubnau, D. (2015). ComGA-RelA interaction and persistence in the *Bacillus subtilis* K-State. *Mol. Microbiol.* 97, 454–471. doi: 10.1111/mmi.13040
- Handke, L. D., Shivers, R. P., and Sonenshein, A. L. (2008). Interaction of *Bacillus subtilis* CodY with GTP. *J. Bacteriol.* 190, 798–806. doi: 10.1128/JB.01115-07
- Harms, A., Maisonneuve, E., and Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* 354:aaf4268. doi: 10.1126/science.aaf4268
- Huynh, T. N., and Woodward, J. J. (2016). Too much of a good thing: regulated depletion of c-di-AMP in the bacterial cytoplasm. *Curr. Op. Microbiol.* 30, 22–29. doi: 10.1016/j.mib.2015.12.007
- Khoo, S. K., Loll, B., Chan, W. T., Shoeman, R. L., Ngoo, L., Yeo, C. C., et al. (2007). Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen *Streptococcus pneumoniae*. *J. Biol. Chem.* 282, 19606–19618. doi: 10.1074/jbc.M701703200
- Korch, S. B., Henderson, T. A., and Hill, T. M. (2003). Characterization of the hipA7 allele of *Escherichia coli* and evidence that high persistence is governed by (p)ppGpp synthesis. *Mol. Microbiol.* 50, 1199–1213. doi: 10.1046/j.1365-2958.2003.03779.x
- Krasny, L., and Gourse, R. L. (2004). An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. *EMBO J.* 23, 4473–4483. doi: 10.1038/sj.emboj.7600423
- Kriel, A., Bittner, A. N., Kim, S. H., Liu, K., Tehrani, A. K., Zou, W. Y., et al. (2012). Direct regulation of GTP homeostasis by (p)ppGpp: a critical component of viability and stress resistance. *Mol. Cell* 48, 231–241. doi: 10.1016/j.molcel.2012.08.009
- Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Dreze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525. doi: 10.1093/nar/gkr131
- Lewis, K. (2010). Persister cells. *Annu. Rev. Microbiol.* 64, 357–372. doi: 10.1146/annurev.micro.112408.134306
- Lioy, V. S., Machon, C., Tabone, M., Gonzalez-Pastor, J. E., Dauglavicius, R., Ayora, S., et al. (2012). The  $\zeta$  toxin induces a set of protective responses and dormancy. *PLoS ONE* 7:e30282. doi: 10.1371/journal.pone.0030282
- Lioy, V. S., Martin, M. T., Camacho, A. G., Lurz, R., Antelmann, H., Hecker, M., et al. (2006). pSM19035-encoded  $\zeta$  toxin induces stasis followed by death in a subpopulation of cells. *Microbiology* 152, 2365–2379. doi: 10.1099/mic.0.28950-0
- Lioy, V. S., Rey, O., Balsa, D., Pellicer, T., and Alonso, J. C. (2010). A toxin-antitoxin module as a target for antimicrobial development. *Plasmid* 63, 31–39. doi: 10.1016/j.plasmid.2009.09.005
- Liu, K., Bittner, A. N., and Wang, J. D. (2015). Diversity in (p)ppGpp metabolism and effectors. *Curr. Op. Microbiol.* 24, 72–79. doi: 10.1016/j.mib.2015.01.012
- Lopez, J. M., Dromerick, A., and Freese, E. (1981). Response of guanosine 5'-triphosphate concentration to nutritional changes and its significance for *Bacillus subtilis* sporulation. *J. Bacteriol.* 146, 605–613.
- Lopez, J. M., Marks, C. L., and Freese, E. (1979). The decrease of guanine nucleotides initiates sporulation of *Bacillus subtilis*. *Biochim. Biophys. Acta* 587, 238–252. doi: 10.1016/0304-4165(79)90357-X
- Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2013). (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* 154, 1140–1150. doi: 10.1016/j.cell.2013.07.048
- Maisonneuve, E., Shakespeare, L. J., Jorgensen, M. G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13206–13211. doi: 10.1073/pnas.1100186108
- Meinhart, A., Alings, C., Strater, N., Camacho, A. G., Alonso, J. C., and Saenger, W. (2001). Crystallization and preliminary X-ray diffraction studies of the e $\zeta$  addition system encoded by *Streptococcus pyogenes* plasmid pSM19035. *Acta Crystallogr. D Biol. Crystallogr.* 57, 745–747. doi: 10.1107/S0907444901004176
- Meinhart, A., Alonso, J. C., Strater, N., and Saenger, W. (2003). Crystal structure of the plasmid maintenance system e $\zeta$ : functional mechanism of toxin  $\zeta$  and inactivation by e $\zeta_2$  complex formation. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1661–1666. doi: 10.1073/pnas.0434325100
- Mutschler, H., Gebhardt, M., Shoeman, R. L., and Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol.* 9:e1001033. doi: 10.1371/journal.pbio.1001033
- Mutschler, H., and Meinhart, A. (2013). Type II Toxin-Antitoxin Loci: The e $\zeta$  Family,” in *Prokaryotic Toxin-Antitoxins*, ed K. Gerdes (Berlin; Heidelberg: Springer-Verlag), 205–223.
- Nanamiya, H., Kasai, K., Nozawa, A., Yun, C. S., Narisawa, T., Murakami, K., et al. (2008). Identification and functional analysis of novel (p)ppGpp synthetase genes in *Bacillus subtilis*. *Mol. Microbiol.* 67, 291–304. doi: 10.1111/j.1365-2958.2007.06018.x
- Nguyen, D., Joshi-Datar, A., Lepine, F., Bauerle, E., Olakanmi, O., Beer, K., et al. (2011). Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science* 334, 982–986. doi: 10.1126/science.1211037
- Oppenheimer-Shaanan, Y., Wexselblatt, E., Katzhendler, J., Yavin, E., and Ben-Yehuda, S. (2011). c-di-AMP reports DNA integrity during sporulation in *Bacillus subtilis*. *EMBO Rep.* 12, 594–601. doi: 10.1038/embor.2011.77
- Pedley, A. M., and Benkovic, S. J. (2017). A new view into the regulation of purine metabolism: the purinosome. *Trends Biochem. Sci.* 42, 141–154. doi: 10.1016/j.tibs.2016.09.009
- Potrykus, K., and Cashel, M. (2008). (p)ppGpp: still magical? *Annu. Rev. Microbiol.* 62, 35–51. doi: 10.1146/annurev.micro.62.081307.162903
- Rao, F., See, R. Y., Zhang, D., Toh, D. C., Ji, Q., and Liang, Z. X. (2010). YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. *J. Biol. Chem.* 285, 473–482. doi: 10.1074/jbc.M109.040238
- Ratnayake-Lecamwasam, M., Serron, P., Wong, K. W., and Sonenshein, A. L. (2001). *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev.* 15, 1093–1103. doi: 10.1101/gad.874201
- Shan, Y., Brown Gandt, A., Rowe, S. E., Deisinger, J. P., Conlon, B. P., and Lewis, K. (2017). ATP-dependent persister formation in *Escherichia coli*. *MBio* 8, e02267–e02216. doi: 10.1128/mbio.02267-16
- Soldo, B., Scotti, C., Karamata, D., and Lazarevic, V. (2003). The *Bacillus subtilis* Gne (GneA, GalE) protein can catalyze UDP-glucose as

- well as UDP-N-acetylglucosamine 4-epimerisation. *Gene* 319, 65–69. doi: 10.1016/S0378-1119(03)00793-5
- Sonenshein, A. L. (2007). Control of key metabolic intersections in *Bacillus subtilis*. *Nat. Rev. Microbiol.* 5, 917–927. doi: 10.1038/nrmicro1772
- Srivatsan, A., and Wang, J. D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr. Op. Microbiol.* 11, 100–105. doi: 10.1016/j.mib.2008.02.001
- Tabone, M., Ayora, S., and Alonso, J. C. (2014a). Toxin  $\zeta$  reversible induces dormancy and reduces the UDP-N-acetylglucosamine pool as one of the protective responses to cope with stress. *Toxins (Basel)* 6, 2787–2803. doi: 10.3390/toxins6092787
- Tabone, M., Lioy, V. S., Ayora, S., Machon, C., and Alonso, J. C. (2014b). Role of toxin  $\zeta$  and starvation responses in the sensitivity to antimicrobials. *PLoS ONE* 9:e86615. doi: 10.1371/journal.pone.0086615
- Unterholzner, S. J., Poppenberger, B., and Rozhon, W. (2013). Toxin-antitoxin systems: biology, identification, and application. *Mob. Genet. Elements* 3:e26219. doi: 10.4161/mge.26219
- Wang, J. D., Sanders, G. M., and Grossman, A. D. (2007). Nutritional control of elongation of DNA replication by (p)ppGpp. *Cell* 128, 865–875. doi: 10.1016/j.cell.2006.12.043
- Witte, G., Hartung, S., Buttner, K., and Hopfner, K. P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* 30, 167–178. doi: 10.1016/j.molcel.2008.02.020
- Wu, X., Wang, X., Drlica, K., and Zhao, X. (2011). A toxin-antitoxin module in *Bacillus subtilis* can both mitigate and amplify effects of lethal stress. *PLoS ONE* 6:e23909. doi: 10.1371/journal.pone.0023909
- Xiao, H., Kalman, M., Ikehara, K., Zemel, S., Glaser, G., and Cashel, M. (1991). Residual guanosine 3',5'-bispyrophosphate synthetic activity of *relA* null mutants can be eliminated by *spoT* null mutations. *J. Biol. Chem.* 266, 5980–5990.
- Yadav, T., Carrasco, B., Myers, A. R., George, N. P., Keck, J. L., and Alonso, J. C. (2012). Genetic recombination in *Bacillus subtilis*: a division of labor between two single-strand DNA-binding proteins. *Nucleic Acids Res.* 40, 5546–5559. doi: 10.1093/nar/gks173
- Yadav, T., Carrasco, B., Serrano, E., and Alonso, J. C. (2014). Roles of *Bacillus subtilis* DprA and SsbA in RecA-mediated genetic recombination. *J. Biol. Chem.* 289, 27640–27652. doi: 10.1074/jbc.M114.577924
- Yamaguchi, Y., Park, J. H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet.* 45, 61–79. doi: 10.1146/annurev-genet-110410-132412
- Zhu, Y., Pham, T. H., Nhiep, T. H., Vu, N. M., Marcellin, E., Chakraborti, A., et al. (2016). Cyclic-di-AMP synthesis by the diadenylate cyclase CdaA is modulated by the peptidoglycan biosynthesis enzyme GlmM in *Lactococcus lactis*. *Mol. Microbiol.* 99, 1015–1027. doi: 10.1111/mmi.13281
- Zielenkiewicz, U., and Ceglowski, P. (2005). The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J. Bacteriol.* 187, 6094–6105. doi: 10.1128/JB.187.17.6094-6105.2005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RDO and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Moreno-del Álamo, Tabone, Lioy and Alonso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Molecular Basis of Stationary Phase Survival and Applications

Jananee Jaishankar and Preeti Srivastava\*

Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, New Delhi, India

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
United States

### Reviewed by:

Grzegorz Węgrzyn,  
University of Gdansk, Poland  
Susana Brom,  
National Autonomous University  
of Mexico, Mexico  
Jan Nesvera,  
Institute of Microbiology of the Czech  
Academy of Sciences, Czechia

### \*Correspondence:

Preeti Srivastava  
preeti@dbeb.iitd.ac.in;  
preetisrivastava@hotmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 31 July 2017

**Accepted:** 28 September 2017

**Published:** 16 October 2017

### Citation:

Jaishankar J and Srivastava P (2017)  
Molecular Basis of Stationary Phase  
Survival and Applications.  
Front. Microbiol. 8:2000.  
doi: 10.3389/fmicb.2017.02000

Stationary phase is the stage when growth ceases but cells remain metabolically active. Several physical and molecular changes take place during this stage that makes them interesting to explore. The characteristic proteins synthesized in the stationary phase are indispensable as they confer viability to the bacteria. Detailed knowledge of these proteins and the genes synthesizing them is required to understand the survival in such nutrient deprived conditions. The promoters, which drive the expression of these genes, are called stationary phase promoters. These promoters exhibit increased activity in the stationary phase and less or no activity in the exponential phase. The vectors constructed based on these promoters are ideal for large-scale protein production due to the absence of any external inducers. A number of recombinant protein production systems have been developed using these promoters. This review describes the stationary phase survival of bacteria, the promoters involved, their importance, regulation, and applications.

**Keywords:** stationary phase promoters, stationary phase gene expression, plasmid vectors, sigma factor, stationary phase

## INTRODUCTION

The majority of the microorganisms around us in air, sea water, and soil are predominantly present in stationary phase (Gefen et al., 2014). The natural habitat of microorganisms often contains limited nutrients due to which rapid growth is usually hampered. Apart from nutrient deprivation, there are other conditions, including physical and chemical stresses, which result in unbalanced growth. All these events result in many changes at the molecular level. These molecular changes are comparable to those observed during the stationary phase of bacteria as witnessed in laboratory studies. The entry of bacteria to the stationary phase can be caused by different factors, including limitation of a specific essential nutrient, accumulation of toxic by-products, presence of stress factors such as changes in pH, temperature, osmolarity, etc. As the cell enters this phase, there is a reduction in cell size and the DNA/protein ratio is said to increase during transition to stationary phase (Nystrom, 2004). The stationary phase has received much attention due to the pattern of protein synthesis in this phase and also because of survival strategies adopted by bacteria. Numerous physiological, morphological, and gene expression changes are observed when a growing cell enters the stationary phase. These are discussed in the following sections.

## PHYSIOLOGY OF THE STATIONARY PHASE

In the stationary phase, the cells become spherical and smaller with a rigid cell envelope, the cell wall is highly cross-linked, membrane fluidity reduces, and cells activate the stringent response mechanism in order to survive the calamity. The activation of this mechanism allows the bacteria to reprogram the gene expression pattern to adapt to different stresses. Two key components of the bacterial stringent response are ppGpp and pppGpp (which are explained in a later section). As a consequence, the cells divert their resources away from growth toward synthesizing amino acids so as to promote survival till nutrient conditions improve.

**Figure 1** depicts the various changes observed in a cell when it enters the stationary phase. The peptidoglycan layer, being the stress-bearing component of cell, increases in thickness. It accounts for 0.7–0.8% of cell's dry weight in exponential phase cells whereas in stationary phase it increases up to 1.4–1.9% (Mengin-Lecreux and van Heijenoort, 1985). At the subcellular level, nucleoid condensation occurs for DNA protection, the cytoplasm gets condensed with an overall decrease in protein synthesis as a consequence of stress or stationary phase (Navarro Llorens et al., 2010). At the translational level, the 70S ribosomes are converted into inactive 100S ribosome dimers by associating with ribosome modulation factor (Wada, 1998). This process, termed as ribosome hibernation, is thought to be a mechanism to fine-tune the translation process according to environmental conditions (McKay and Portnoy, 2015). Recently, 16S rRNA fragmentation at the tip of helix 6 has been shown to attenuate the activity of 30S ribosomal subunit and thereby protein synthesis (Luidalepp et al., 2016). Also, during limited nutrient availability, accumulation of truncated mRNA and deacylated tRNA occurs. The ribosomes become stuck on these mRNAs and owing to the absence of a stop codon, the ribosome is unable to get released (Pletnev et al., 2015). These mechanisms are understood to be the defense response upon starvation. As a result of the various morphological, metabolic, transcriptional, or translational alterations, the stationary phase cells become resistant to high temperature, high concentrations of H<sub>2</sub>O<sub>2</sub>, and very high medium osmolality.

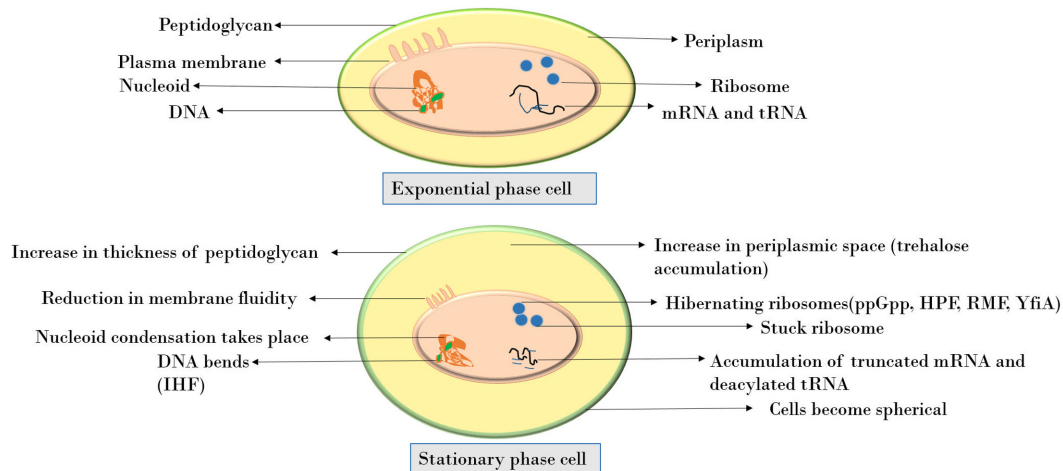
Cells in exponential, stationary, and long-term stationary phases have different fates (**Figure 2**). As a consequence of starvation, many bacteria including the genera *Bacillus* and *Clostridium* form resistant spores helping them withstand the harsh surrounding environment. Non-optimal growth conditions also lead to the formation of biofilm in many bacterial species. Physiologically, biofilm bacteria are similar to stationary phase bacteria. One key transition is the formation of persisters induced during stationary phase, in biofilms, and also as a consequence of a general stress response. These cells could also arise in exponential growth by the activation of ppGpp as a consequence of sub-lethal antibiotic concentration. The formation of these bacterial persisters is understood to be the reason behind relapsing infections and is a major cause of drug resistance (Harms et al., 2016).

During the late stationary phase sometimes referred to as long-term stationary phase, several remarkable adaptations take place. On continued starvation, one of the survival strategies includes bacteria entering a viable but non-culturable state (VBNC). In this state, bacteria remain metabolically active but fail to form colonies on bacteriological media. Several bacteria including *Rhodococcus biphenylivorans* (Su et al., 2015), *Escherichia coli*, *Agrobacterium tumefaciens*, *Helicobacter pylori*, *Lactococcus lactis*, many *Vibrio* species, and *Pseudomonas* species have been shown to enter the VBNC state (Oliver, 2005). The VBNC state poses a serious health risk as the dormant bacterial species could remain undetected in culturable conditions, though having the ability to cause infections (Navarro Llorens et al., 2010). A variety of stresses is said to lead to the manifestation of VBNC state (Pletnev et al., 2015). Prolonged starvation also results in Growth Advantage in Stationary Phase (GASP) phenotype. The GASP phenomenon is a result of mutations in the *rpoS* allele (described later) which confers a gainful ability to continue growing during starvation conditions, thus replacing the parental population (Navarro Llorens et al., 2010). These mutations allow the mutants to effectively scavenge the nutrients released by dead cells (Zambrano and Kolter, 1996). A number of Gram-positive bacteria such as *Listeria monocytogenes* (Bruno and Freitag, 2011), *Staphylococcus aureus*, *Enterococcus faecalis*, and *Bacillus globigii* (Finkel et al., 1997) and Gram-negative bacteria including *Campylobacter*, *Geobacter*, *Vibrio*, *E. coli*, *Pseudomonas*, etc., have been found to enter the GASP state (Chen and Chen, 2014). Gefen et al. (2014) coined the term 'constant activity stationary phase' (CASP) to describe the phenomenon of constant rate of protein synthesis observed in non-growing bacteria that have undergone over more than 60 h of starvation. On studying the protein production at this stage, they have found that both the protein synthesis machinery including ribosomes, RNA polymerases, etc., and resources such as amino acids, nucleotides, etc., remain constant at CASP. Finally, constant promoter activity was observed in this experiment for up to 10 h of starvation. Another interesting phenomenon experienced by bacterial population in stationary phase is the 'stationary phase contact-dependent inhibition' (SCDI). It requires physical contact between the evolved and original bacteria (Lemonnier et al., 2008). In this process, it was observed that the evolved strains either killed or inhibited the growth of bacteria that they were derived from. The inhibiting ability of these strains is attributed to mutations within a single gene involved in glycogen synthesis pathway: *glgC* (encoding ADP-glucose pyrophosphorylase). Astonishingly, all evolved strains overproduced glycogen which seemed to be necessary for SCDI to occur (Navarro Llorens et al., 2010).

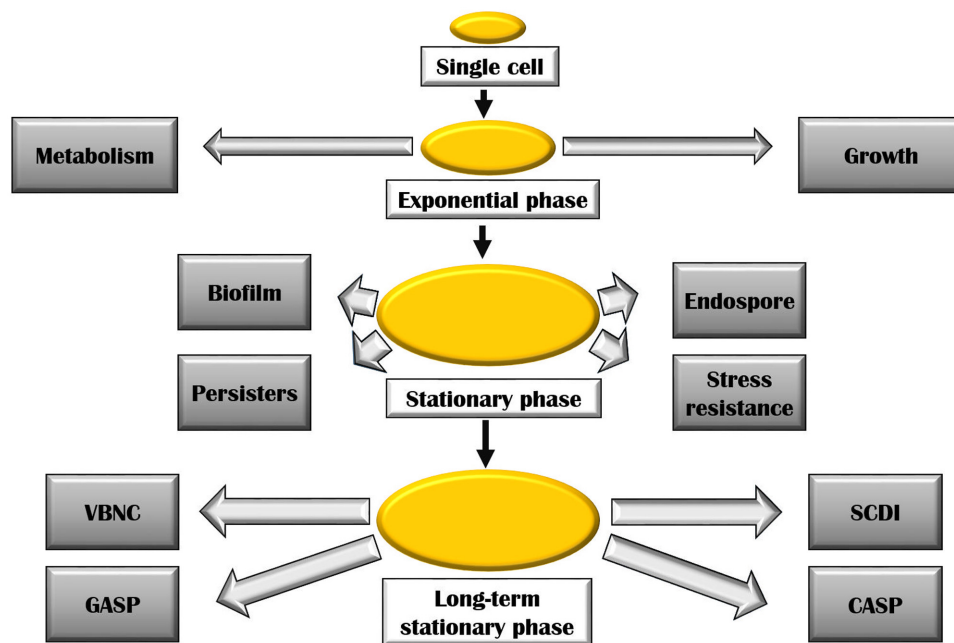
## ALTERNATIVE SIGMA FACTORS ACTIVE AT STATIONARY PHASE

A key regulator of stationary phase gene expression in *E. coli* is the transcription factor  $\sigma^S$  [a product of *rpoS* (*katF*) gene]. The *E. coli* genome was found to contain two genes *katE* and *katG* encoding for HPII and HP1w1-4x catalases. The expression of HPII was





**FIGURE 1** | Comparison of molecular and cellular changes in exponential vs stationary phase.



**FIGURE 2** | Various bacterial adaptations at stationary and long-term stationary phase. Abbreviations are described in the text.

highest in stationary phase and has been shown to be completely dependent on *katF* gene product. The latter serves as sigma factor for RNA polymerase and therefore named as *rpoS* or  $\sigma^S$  or  $\sigma^{38}$  or stationary phase sigma factor or starvation sigma factor (Tanaka et al., 1997).

The amount of  $\sigma^S$  remains relatively low in the growing phase of cells but increases markedly when the cell encounters stress, starvation or enters stationary phase. The role of this protein is to aid in survival and improved resistance to stressful conditions. Induction of  $\sigma^S$  is observed under conditions of low pH, heat or cold shock, UV-induced DNA damage, nutrient starvation,

high cell density, high osmolarity, etc. (Hengge, 2011). The  $\sigma^S$ -dependent genes have been attributed to morphological changes (Hengge, 2011), induction of starvation proteins (Alexander and St. John, 1994), iron uptake, carbohydrate metabolism, amino acid transport, and so on, at the onset of stationary phase (Lacour and Landini, 2004).

The *rpoS* sigma factor is selectively utilized in stationary phase. The major sigma factor *rpoD* ( $\sigma^{70}$ ) is inhibited by a regulator of sigma D (Rsd). The rationale for  $\sigma^S$  selectivity *in vivo* is not completely understood, but it is known that many promoters can exhibit both  $\sigma^S$  and  $\sigma^{70}$  mediated expression

*in vitro*. It is well known that  $\sigma^{70}$  is affected by changes in spacer region and consensus -10 and -35 positions, but the alternative  $\sigma^S$  is shown to be less affected by changes in these regions, thus making it more selective *in vivo* (Hengge, 2011). Another observation by Tanaka et al., 1995 indicates that the -35 region is not always required for stationary-phase expression (Tanaka et al., 1995). In this study, the *fic* promoter was shown to function with promoter sequences downstream from -17. Also, the promoters recognized by RpoS are found to contain curved DNA region. Hence, the absence of consensus -35 and the presence of curved DNA region imparted  $\sigma^S$  dependence to galP1 and galP2 promoters, whereas the presence of -35 sequence in the same promoter changed the specificity toward  $\sigma^{70}$  (Kolb et al., 1995). Thus, the general belief is that the  $\sigma^S$  promoters lack a -35 consensus sequence. However, some authors have suggested CTGCAA (Bohannon et al., 1991) or CCGACA (Wise et al., 1996) as the -35 consensus sequence. Similarly for -10, Hengge-Aronis (1993) has suggested a consensus sequence of TATACT, which was later changed to CTATACT (Espinosa-Urgel et al., 1996). More recently, a long consensus sequence KCTAYRCTTAA for -10 region has been proposed, where K could be T or G, Y could be T or C, and R could be A or G (Weber et al., 2005). Not all the stationary-phase induced genes depend on  $\sigma^S$ , and out of the many genes that show higher level of expression in the stationary phase, only 10% is known to be dependent on  $\sigma^S$  (Rava et al., 1999). Out of the genes induced in stationary phase, those that show  $\sigma^S$  independent behavior are *dnaK*, *groEL*, *htpG* which depend on  $\sigma^{32}$  (Kolter et al., 1993).

Several other alternative sigma factors have been reported. In *Salmonella typhimurium*,  $\sigma^E$  has been suggested to serve a complementary role in stationary phase survival. Mutants deficient in *rpoH* gene coding for  $\sigma^E$  have been shown to be susceptible to oxidative stress (Testerman et al., 2002).

The number of sigma factors varies from 1 in *Mycoplasma genitalium* (Dorman, 2011), 6 in *Gordonia* sp. IITR100 (Jaishankar et al., 2017), 7 in *E. coli* (Ishihama, 1997), 18 in *B. subtilis* (Gruber and Gross, 2003), 24 in *Pseudomonas aeruginosa* (Potvin et al., 2008), and 65 in *Streptomyces coelicolor* (Kim et al., 2008). **Table 1** gives a list of various sigma factors in well-known bacterial species and the types of sigma factors upregulated at stationary phase.

## REGULATION OF RpoS

The RpoS is regulated at post-transcriptional level by *rpoS* mRNA secondary structure, small RNAs, Hfq, and HU proteins, ClpXP protease and RssB (phosphorylation-modulated RpoS recognition factor) (Hengge-Aronis, 2002). The *rpoS* mRNA is stimulated by regulatory factors such as Hfq (HF-1) protein and DsrA (small regulatory RNA) and repressed by H-NS (histone-like protein) and *oxyS* RNA. The 5' UTR of *rpoS* mRNA forms a loop which represses its translation. This loop can be disrupted by non-coding RNAs such as *dsrA*, *rprA*, and *arcA* (Gaida et al., 2013). Another sRNA which positively regulates *rpoS* mRNA is *gcvB* (Jin et al., 2009).

The turnover of RpoS protein in exponential phase is very high with a half-life of 1.4 min (Lange and Hengge-Aronis, 1994). The RpoS protein is stable in stationary phase.

The levels of RpoS are also controlled by a number of other factors. These include both positive regulators such as ppGpp and polyphosphate (polyP) and negative regulators such as cAMP and UDP glucose.

The availability of ppGpp is dependent on RelA, a ppGpp synthase that is associated with ribosomes. In stationary phase, when the uncharged tRNAs accumulate due to decreased availability of amino acids, *relA* is turned on and synthesizes ppGpp. This turns on the promoters involved in amino acid biosynthesis and uptake (Barker et al., 2001). It has been shown that 6S RNA regulates *relA* gene expression, which leads to alteration in ppGpp levels in stationary phase (Cavanagh et al., 2010). The rRNA genes are turned off by ppGpp. Many stationary phase promoters (SPPs) are also regulated by 6S RNA, even in the absence of ppGpp.

In *B. subtilis*, it has been demonstrated that cells entering in stationary phase have small GTP and GDP pools. This is possibly due to conversion of GTP to (p)ppGpp or due to the lack of sufficient precursors available for nucleotide synthesis. Lopez and coworkers demonstrated that treatment of cells with decoyinine, an inhibitor of GMP synthase, can result in induction of stationary phase genes (Ratnayake-Lecamwasam et al., 2001).

The intracellular levels of certain compounds such as trehalose, glycine betaine, glycogen, and polyphosphate are high under stress conditions. Some of these compounds modulate function of the RpoS holoenzyme. For example, glutamate and trehalose modulate the holoenzyme binding to promoters. Similarly, altered promoter selectivity has been observed in *E. coli* when RpoS associates with inorganic polyphosphate. The inhibition due to PolyP is relieved by high concentrations of potassium glutamate (Shimada et al., 2004). Bacterial pheromone, Homoserine lactone (HSL), a small molecule responsible for communication between bacteria, also affects the concentration of  $\sigma^S$  in the cell. Mutants in the biosynthetic pathway for synthesis of HSL lose the ability to induce  $\sigma^S$  (Zambrano and Kolter, 1996).

## EXPRESSION OF GENES IN STATIONARY PHASE

When the cells are growing, the metabolism-linked genes are highly expressed, and get turned off when the cells enter stationary phase. The stationary phase is a period of no growth, however, genes essential for survival of organisms are expressed at this stage. Around 20% of the genes of *E. coli* are found to express at higher level in the stationary phase (Rava et al., 1999). These genes are directly linked to many key events including DNA repair, glycogen production, thermotolerance, osmotolerance, etc. (Bohannon et al., 1991; Ishihama, 1997). Transcriptome profiling/expression analysis of *E. coli* in stationary phase revealed upregulation of genes which are involved in survival during osmotic stress (*ots*, *tre*, *osm*), long-term survival (e.g., *bolA*, *dps*, *cbpA*, and *glgS*), periplasmic shock (*rpoE* and *rseA*),

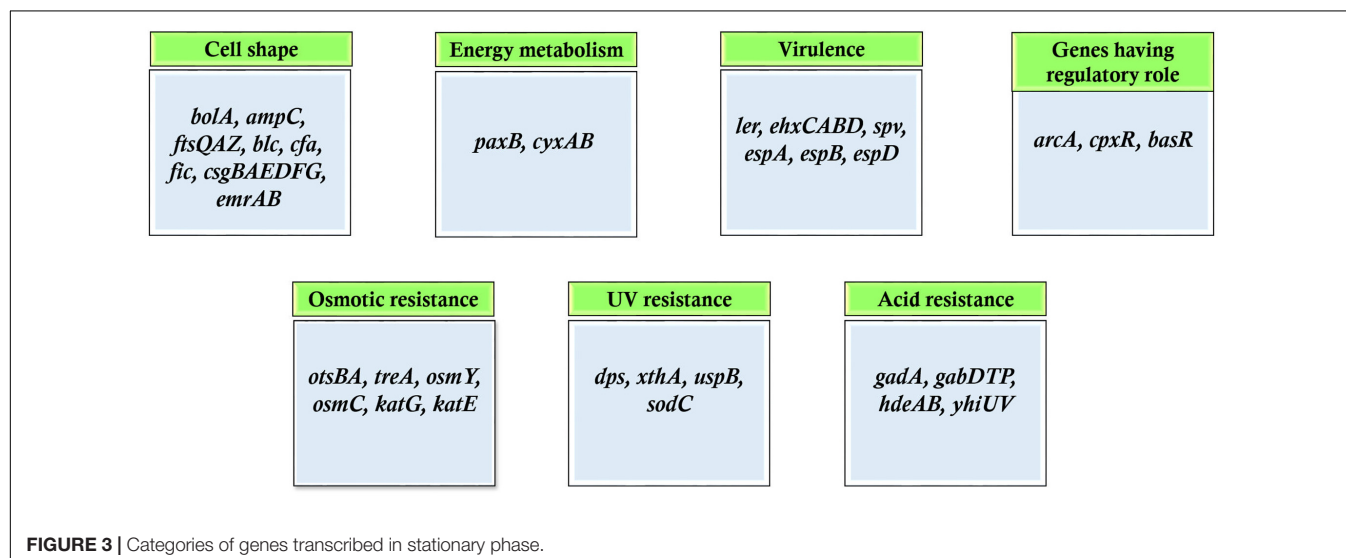
**TABLE 1** | List of sigma factors upregulated at stationary phase in different bacteria.

Name of the organism	Sigma factors	Sigma factors upregulated at stationary phase	Reference
<i>Escherichia coli</i>	<b>7</b> $\sigma^{70}(\sigma^D)$ , $\sigma^{24}(\sigma^E)$ , $\sigma^{28}(\sigma^F)$ , $\sigma^{32}(\sigma^H)$ , $\sigma^{54}(\sigma^N)$ , $\sigma^{38}(\sigma^S)$ , $\sigma^{18}(\sigma^{FecI})$	$\sigma^{32}(\sigma^H)$ , $\sigma^{54}(\sigma^N)$ , $\sigma^{38}(\sigma^S)$	Ishihama, 1997
<i>Bacillus subtilis</i>	<b>18</b> $\sigma^A$ , $\sigma^B$ , $\sigma^C$ , $\sigma^D$ , $\sigma^E$ , $\sigma^F$ , $\sigma^G$ , $\sigma^H$ , $\sigma^K$ , $\sigma^L$ , $\sigma^M$ , $\sigma^V$ , $\sigma^W$ , $\sigma^Y$ , $\sigma^X$ , $\sigma^Z$ , Xpf, YlacC,	$\sigma^B$ , $\sigma^C$ , $\sigma^D$ , $\sigma^H$	Haldenwang, 1995; Gruber and Gross, 2003
<i>Streptomyces coelicolor</i>	<b>65</b>	$\sigma^B$ , $\sigma^F$ , $\sigma^H$ , $\sigma^M$ , $\sigma^N$ , $\sigma^R$	Kim et al., 2008; Tripathi et al., 2014
<i>Pseudomonas aeruginosa</i>	<b>24</b>	$\sigma^E$ , $\sigma^H$ , $\sigma^S$	Potvin et al., 2008
<i>Corynebacterium glutamicum</i>	<b>7</b> $\sigma^A$ , $\sigma^B$ , $\sigma^C$ , $\sigma^D$ , $\sigma^E$ , $\sigma^H$ , $\sigma^M$	$\sigma^B$ , $\sigma^H$ , $\sigma^M$	Pátek and Nešvera, 2013

cold shock (*csp* genes), etc. Other genes include carbon storage regulator (*csrA*), trp repressor binding protein (*wrbA*) and universal stress protein (*uspA*) (Chang et al., 2002). Moreover, several antibiotics including lactocin B of lactic acid bacteria, alfatoxin of *Aspergillus* species are produced mainly in stationary phase (Matin, 1992).

Persister cell formation has also been attributed to genes differentially expressed in stationary phase. These cells are recalcitrant to antibiotic treatments and often are the major cause of drug resistance. Several polyamines including putrescine, spermidine, and cadaverine direct persister formation through upregulation of genes such as *rpoS*, *rmf*, *yqjD* (Tkachenko et al., 2017). This observation suggests that polyamine metabolism participates in the regulation of persister cells formation. To determine the genes upregulated at stationary phase microarray was done in *Mycobacterium smegmatis* grown under conditions of glycerol and glucose depletion. Different subset of genes were identified that were preferentially upregulated at stationary phase. The categories of genes included those involved in metabolism of sulfur, sigma factors including *sigB*, *sigE*, and *sigH*, fatty

acid degradation, anaerobic respiration, etc. (Hampshire et al., 2004). Also, of key interest in this study is the presence of stationary phase operons involving many gene clusters that were significantly upregulated in stationary phase. On investigating further, the presence of other such operons were also found. The *pdh* operon of *Streptococcus mutans* is expressed only in the stationary phase. This operon was observed to be transcribed only by a subpopulation of bacteria in stationary phase and was vital for survival during long periods of sugar starvation. The *pdh* operon consists of four genes that are transcribed as an operon: *pdhD*, *pdhA*, *pdhB*, *pdhC*, which encode the components of PDH (pyruvate dehydrogenase) complex, i.e., pyruvate dehydrogenase (two subunits encoded by *pdhA* and *pdhB*), dihydrolipoyl transacetylase (*pdhC*), and dihydrolipoyl dehydrogenase (*pdhD*). The inactivation of the first gene: *pdhD* resulted in impaired survival in both batch cultures and biofilms (Busuioc et al., 2010). Similarly, phage shock protein operon (*pspABCE*) of *E. coli* was reported to be critical for survival under prolonged stationary phase at alkaline conditions. This operon was expressed strongly



**TABLE 2 |** Stationary phase promoters in Gram-negative bacteria.

Name of or	Name of promoter	Gene product	–10	–35	Other motifs	Length of spacer	Reference
<i>Escherichia coli</i>	<i>BolAp1</i>	BolA	CGGCTAGTA	CTGCAA	–	15	Aldea et al., 1989
	<i>TreA</i>	TreA (Osmotically inducible periplasmic trehalase)	ATGCAG	TAAGGT	–	17	Repoila and Gutierrez, 1991
	<i>Cst-1</i>	Cst	–	–	–	–	Tunner et al., 1992
	<i>Fic</i>	Fic (PABA or folate)	TATACT	–	–	–	Utsumi et al., 1993
	<i>Hns</i>	H-NS	TATTAT	TTGCAC	–	17	Dersch et al., 1993
	<i>PoxB</i>	PoxB (pyruvate oxidase)	TAAACT	–	–25: CGTCA; –60: GTTAGTG	–	Chang et al., 1994
	<i>Slp</i>	Slp	TATTATG	GATGAAA	–	16	Alexander and St. John, 1994
	<i>AldB</i>	AldB (Aldehyde dehydrogenase)	TACCCT	–	–	–	Xu and Johnson, 1995
	<i>CsiD</i>	CsiD (Carbon starvation inducible gene)	TATTTT	TGCGCA	–	17	Marschall et al., 1998
	<i>OsmY (Csi-5)</i>	OsmY (Periplasmic protein of unknown function)	TATATT	CGAGCG	–	15	Lange et al., 1993; Becker and Hengge-Aronis, 2004
<i>Shigella flexneri</i>	<i>GadA</i>	GadA (Glutamate decarboxylase)	CTACTTT	–	–	–	Waterman and Small, 2003
<i>Vibrio anguillarum</i>	<i>EmpA</i>	EmpA metalloprotease	GATCCA	CCGTGCTAC	–	19	Croxatto et al., 2004; Denkin and Nelson, 2004

**TABLE 3 |** Stationary phase promoters from Gram-positive bacteria.

Organism	Promoter	Gene product	–10	–35	Length of spacer (bp)	Reference
<i>Bacillus subtilis</i>	<i>Pst</i>	Phosphate-specific transport	TTTACT	TTCAAA	18	Qi et al., 1997
<i>Bacillus subtilis</i>	<i>Cry3a</i>	Crystal proteins	TAAGCT	TTGCAA	18	Lee et al., 2010
<i>Bacillus subtilis</i>	<i>Ylb</i>	–	TACAAT	TTGGA	18	Yu et al., 2015
<i>Bacillus subtilis</i>	<i>SrfA</i> mutant	Srf operon (lipopeptide antibiotic surfactin)	TTGACT	TATAAT	–	Guan et al., 2016
<i>Streptomyces coelicolor</i>	<i>KasO</i> mutant	Colemycin P1	TAAAGT	TTGACA	18	Wang et al., 2013
<i>Corynebacterium glutamicum</i>	<i>Cg3141</i> mutant	<i>Cg3141</i> (flavohemoprotein)	TGGGAT	TTAAGG	17	Kim et al., 2016
<i>Gordonia</i> sp. IITR100	Stationary phase promoter	–	AATAAT	TTAACT	22	Singh et al., 2016

under extreme stressful conditions and remained significant for survival under nutrient-limited conditions (Weiner and Model, 1994). Categories of genes that are preferentially upregulated in stationary phase is shown in **Figure 3**. Studies have demonstrated that starved cells exhibit more protective resistance to different stresses as compared to resistance induced during growing stage by non-lethal exposure of stresses (Kolter et al., 1993).

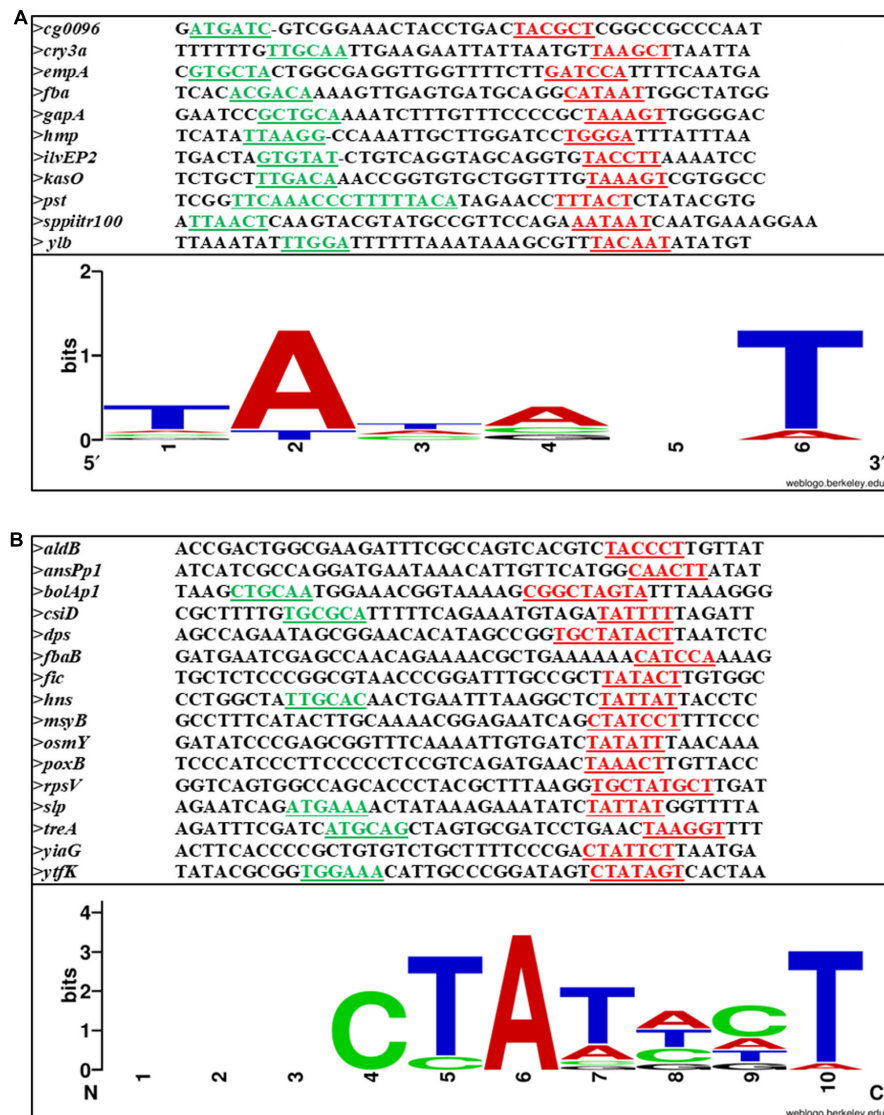
## STATIONARY PHASE PROMOTERS

The genes expressed in stationary phase are controlled by promoters, which result in induction of stationary phase. The promoters, which are turned on, are called SPPs. They are

recognized by RNA polymerase holoenzyme containing  $\sigma^S$  and therefore called RpoS.

The vast importance of SPPs had been realized way back in 1980s with the study of *mcbA* promoter and *bolA* P1 promoter of *E. coli* (Connell et al., 1987; Aldea et al., 1989). The *mcbA* promoter causes increased level of transcription initiation for Microcin B17, a DNA replication inhibitor. Promoter *mcbA*-LacZ fusion showed the induction of transcription in nitrogen, phosphate, and carbon starvation conditions. Similarly, *bolA*-lacZ fusion demonstrated an increase in expression of approximately 10- to 20-fold during transition to stationary phase. Since then many SPPs have been isolated and characterized in both Gram-positive and Gram-negative bacteria (**Tables 2, 3**). Particularly regarding *E. coli* and *B. subtilis*, the stationary-phase-specific





**FIGURE 4 |** Sequence alignment of (A) Gram-positive and (B) Gram-negative stationary phase promoters. -10 and -35 are underlined and shown in red and green, respectively. The conserved bases are shown below.

gene regulation has been intensively studied (Hengge, 2011).

On analysis of the different SPPs, our observation is that there is not much variation between this class of promoters and  $\sigma^{70}$  promoters. It is the sequence outside the -10 and -35 regions that distinguish between  $\sigma^{70}$ - and  $\sigma^S$ -dependent promoters. **Figures 4A,B** shows the -10, -35 and spacer region of few SPPs from Gram-positive and Gram-negative bacteria and the consensus sequence at the -10 region is shown as a logo designed using WebLogo software available online (Crooks et al., 2004).

Among the SPPs exist a special class of promoters known as the gearbox promoters which include *mcbAp*, *bolAp1*, *ftsQp* to name a few. This class of promoters has been studied in several Gram-negative bacteria including *E. coli*. Two different highly

conserved consensus -10 and -35 sequence have been proposed by Aldea et al. (1993) for this class of promoters: CTGCAA or GTTAAGC at -35 position and CGGCAAGTA or CGTCC at -10 position. Gearbox promoter-induced gene expression seems to correlate inversely with growth rate and these promoters may or may not depend on  $\sigma^S$ .

## ENERGY RESERVES CONSUMED DURING STATIONARY PHASE AND SOURCE OF NUTRIENTS FOR PROTEIN PRODUCTION

During unfavorable conditions of growth, reprogramming the cellular machinery for sustaining viability is a natural process

of adaptation. Reserve polymers like glycogen and poly- $\beta$ -hydroxybutyric acid that are accumulated by bacteria during growth are rapidly consumed during conditions of carbon starvation to ensure survival. In case of bacteria that do not accumulate these polymers, cellular RNA is rapidly degraded for energy generation (Matin, 1992). Among RNA, rRNA is preferentially degraded (Deutscher, 2003). Besides, 50% of ribosomes synthesized during exponential growth are degraded during entry to stationary phase (Piir et al., 2011). What is surprising is that, when in stationary phase, these ribosomes are fairly stable and so degradation occurs only in between the stages.

The yield of protein production from stationary phase systems is as high as 121% as compared to their log phase counterparts (Ou et al., 2004). This raises a very important question: What makes protein synthesis possible at stationary phase?

Balaban and coworkers devised a microfluidic device and followed the production of fluorescent proteins at stationary phase. They found that cells after entering stationary phase continue to produce proteins for several days (Gefen et al., 2014). It has been suggested that cells continue to produce proteins at stationary phase by reusing amino acids from degraded proteins. Moreover, the biosynthetic pathway of a few amino acids including serine, aspartate/asparagine, glutamine/glutamate, and alanine were shown to be active during stationary phase (Shaikh et al., 2010). In addition, it is shown that each condition resulting in starvation results in induction of specific set of proteins (Kolter et al., 1993).

## DEVELOPMENT OF GENE EXPRESSION SYSTEMS USING STATIONARY PHASE PROMOTERS

A strong promoter is the key for developing efficient gene expression systems. For recombinant protein production, several bacterial hosts have been used as cell factories, with features such as easy purification, improved protein folding and secretion, high production of membrane proteins, etc. (Ferrer-Mirallès and Villaverde, 2013). To develop more such expression systems in bacteria, it is necessary to ensure proper selection of a promoter that would drive the expression of genes at the right time and with maximum amount.

Promoters could be classified as constitutive or inducible, growth-stage limited, tissue specific, etc. Inducible promoters can further be classified into inducer-specific and auto-inducible promoters. Constitutive promoters are not useful for toxic proteins. Inducer-specific promoters involve the cost of inducer. Also, some chemical inducers such as Isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) are expensive and toxic (Cao and Xian, 2011). Further, the addition of external inducers often requires growth monitoring which is vital for productivity and hence lead to difficulty in fermentation.

Auto-inducible promoters are ideal for large-scale protein production as they are induced at late log phase or stationary phase. Such promoters induce expression of the recombinant gene without any inducer and thus are economical. However, most of them have low activity (Yu et al., 2015). In *B. subtilis*,

Fan and coworkers successfully identified a strong SPP Pylb by microarray approach (Yu et al., 2015). The  $\beta$ -galactosidase activities were observed to be up to 5000 miller units. The authors have proposed that such a promoter will be useful for protein production. A SPP-based auto-inducible gene expression system has been constructed using cry3Aa promoter. The Pcry drives the expression of crystal proteins in *B. thuringiensis*. The promoter cry3Aa was tested in *B. subtilis* and the wild type have the LacZ levels up to 1000 miller units and on mutagenesis resulted in levels up to 5200 miller units (Lee et al., 2010). Similarly, in another Gram-positive bacteria, *Gordonia* sp. IITR100, a SPP was identified and the  $\beta$ -galactosidase activities were up to 600 miller units (Singh et al., 2016). However, the  $\beta$ -galactosidase activities vary with respect to strain, copy number of plasmid, growth medium, temperature, etc., so it is difficult to assess the strength of promoter based on Miller units alone. In future, a study of such promoters based on the number of transcripts would be useful to compare the strength.

In *Corynebacterium glutamicum*, promoter of *cg3141* gene coding for flavohemoprotein was found to show higher inducibility in the stationary phase. Then, a synthetic promoter library was prepared to change the spacer and flanking regions in the promoter, to obtain a range of promoter strengths (Kim et al., 2016). At the end, one of the synthetic promoters that showed up to 20-fold higher strength compared to the original *cg3141* promoter was obtained and demonstrated for fed-batch cultivation of glutathione S-transferase in a 5L reactor. **Table 4** depicts the list of SPP-based expression vectors constructed till date. Studies like these indicate that the potential of SPPs is phenomenal. In *Streptomyces*, a high-level recombinant protein expression system has been patented (US Patent No. 7,316,914).

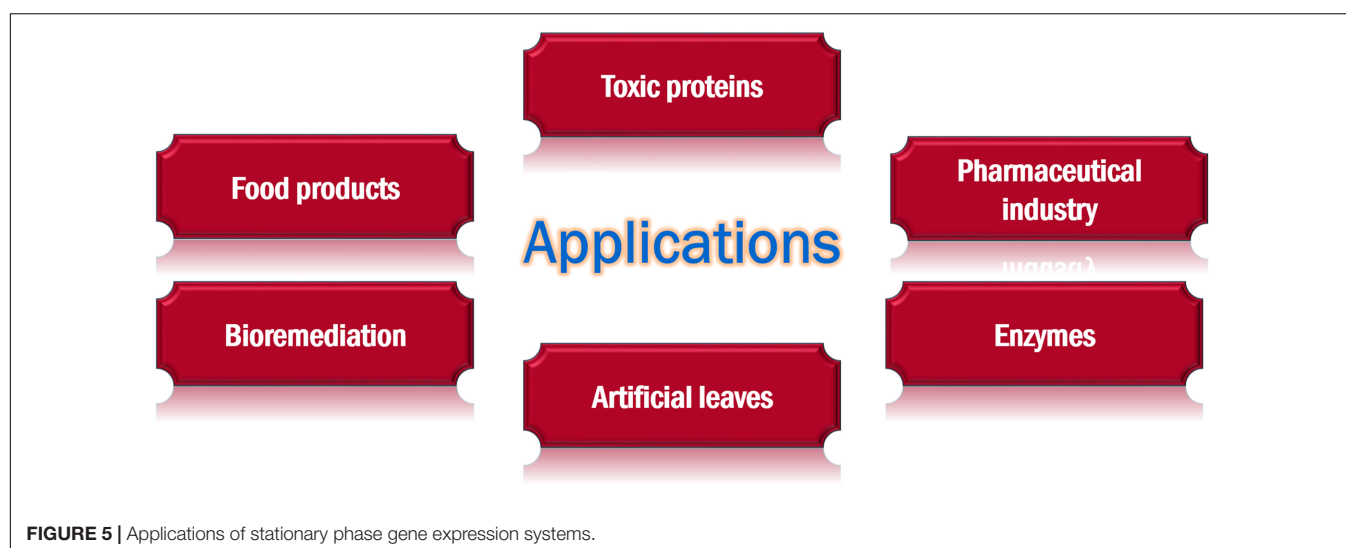
## APPLICATIONS

SPPs have immense potential for use in many industries (**Figure 5**).

Recombinant production of toxins whose overproduction is detrimental to the growth of cells needs controlled conditions for expression. In such cases, the use of SPP is advantageous as the overproduction will not affect the growth of the host cells. Many bacteria have been used to demonstrate the utility of cell-density-dependent expression systems for heterologous protein production. Metabolic engineering of bacteria for enhanced production of industrially important chemicals has been carried out since a long time. The *fic* promoter of *E. coli* was used to express *phlD* gene at a higher titer in stationary phase, without the addition of any inducer, for the production of phloroglucinol, which has utility in pharmaceutical industry and plant tissue culture. After 20 h of cultivation in a flask with shaking, 9% of glucose supplied had converted to phloroglucinol with a productivity of 0.014g/l h (Cao and Xian, 2011). *B. subtilis* has been engineered for overproduction of aminopeptidase using a mutated *Psrfa* system and has resulted in 87.89 U/ml of enzyme activity (Guan et al., 2015). Using *B. subtilis*, a cry-promoter-based system was developed wherein

**TABLE 4 |** Stationary phase promoter-based gene expression systems reported from Gram-negative and Gram-positive bacteria.

Vector	Promoter	Organism	Ori	Antibiotic resistance	Reporter gene	Reference
pTGV1	<i>BolAp</i>	<i>E. coli</i>	pBR322	Ampicillin	<i>bolA-LacZ</i>	Aldea et al., 1989
pFL1	<i>Fic</i>	<i>E. coli</i>	pMB1	Tetracycline	<i>LacZ</i>	Utsumi et al., 1993
pKS4, pKS5	<i>Hns</i>	<i>E. coli</i>	pSC101	Tetracycline	<i>Hns-LacZ</i>	Dersch et al., 1993
pNH5	<i>OsmY (Csi5)</i>	<i>E. coli</i>	pBR322	Ampicillin	<i>Csi5-LacZ</i>	Lange et al., 1993
pMC719	<i>Slp</i>	<i>E. coli</i>	pBR322	Tetracycline	<i>Slp-LacZ</i>	Alexander and St. John, 1994
pYYC128	<i>PoxB</i>	<i>E. coli</i>	pBR322	Chloramphenicol	<i>PoxB-LacZ</i>	Chang et al., 1994
pRJ4025	<i>AldB</i>	<i>E. coli</i>	pBR322	Ampicillin	<i>LacZ</i>	Xu and Johnson, 1995
pYQ23	<i>Pst</i>	<i>Bacillus subtilis</i>	pBR322	Ampicillin, Chloramphenicol	<i>LacZ</i>	Qi et al., 1997
pCM3	<i>CsiD</i>	<i>E. coli</i>	pMB1	Ampicillin	<i>Csi-LacZ</i>	Marschall et al., 1998
pDM35-EmpA	<i>EmpA</i>	<i>Vibrio anguillarum</i>	R6K	Chloramphenicol	<i>LacZ</i>	Croxatto et al., 2004; Denkin and Nelson, 2004
pGRP (many promoters)	Many	<i>E. coli</i>	pBR322	Ampicillin	green fluorescent protein (eGFP)	Shimada et al., 2004
pD82-aprE	<i>Cry3a</i>	<i>Bacillus subtilis</i>	–	Chloramphenicol	<i>AprE-LacZ</i>	Lee et al., 2010
pMD-ficD	<i>Fic</i>	<i>E. coli</i>	pBR322	Ampicillin	<i>phlD</i>	Cao and Xian, 2011
pDR4-K*	<i>KasO</i>	<i>Streptomyces coelicolor</i>	EBV origin	Hygromycin	<i>XylE-neo</i>	Wang et al., 2013
pBSG03	<i>SrfA</i>	<i>Bacillus subtilis</i>	pBR322	Ampicillin, Kanamycin, Neomycin	<i>fp</i>	Guan et al., 2015
Pylb-bgaB-pUBC19	<i>Ylb</i>	<i>Bacillus subtilis</i>	pUBC19	Ampicillin	Beta-gal ( <i>bgaB</i> )	Yu et al., 2015
pSC1	Stationary phase promoter	<i>Gordonia sp.</i> IITR100	pRC4	Kanamycin	<i>LacZ</i>	Singh et al., 2016
pCES-P <sub>4</sub> -N <sub>14</sub> -sfGFP	P <sub>4</sub> -N <sub>14</sub>	<i>Corynebacterium glutamicum</i>	pCG1	Kanamycin	<i>sfGFP</i>	Kim et al., 2016

**FIGURE 5 |** Applications of stationary phase gene expression systems.

cellulose and alkaline protease were produced with a higher yield as compared to the wild-type *cry3A* promoter (Lee et al., 2010).

It is a well-known fact that the non-growing phase of lactic acid bacteria accounts for a major proportion of flavor production in lactic acid bacteria (van de Bunt et al., 2014).

Therefore, engineering bacterial cells in such a way that they are expressed at high levels, during the ripening process, by using SPPs would enhance their applicability in food industry.

In the bioremediation industry, microorganisms have routinely been employed for removing pollutants. Due to low nutrient availability in polluted sites, genetic engineering of cells resulting in higher enzymatic activities at lower growth rates have been shown to be highly efficient for bioremediation process. On studying the phenol degradation capability of two non-growing recombinant *E. coli* strains, it was found that the *groEL*-promoter-driven gene expression system caused 75% phenol degradation while the *tac*-promoter-driven expression could cause only 15% degradation of phenol (Matin, 1992). As suggested by Tunner et al. (1992), it is possible to use starvation-induced promoters for chemical waste biodegradation wherein enzymes can be induced naturally by bacteria due to the occurrence of nutrient-limited conditions in the environment. This could save the cost of induction thereby increasing the efficiency of the process.

In a very interesting experiment, *Rhodospirillum rubrum* cells grown photoheterotrophically, evolved hydrogen for about 70 h after growth ceased (Melnicki et al., 2008). Similarly, a purple non-sulfur photosynthetic bacterium, *Rhodospseudomonas palustris* under nitrogen starvation conditions, produced hydrogen gas for over 4000 h thus paving way for creation of 'artificial leaves' (Gosse et al., 2010).

## CONCLUSION AND FUTURE PROSPECTS

Stationary phase survival is a means of bacterial adaptation by which bacteria survive under conditions of stress or starvation.

## REFERENCES

- Aldea, M., Garrido, T., Hernandez-Chico, C., Vicente, M., and Kushner, S. R. (1989). Induction of a growth-phase-dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene. *EMBO J.* 8, 3923–3931.
- Aldea, M., Garrido, T., and Tormo, A. (1993). Gearbox gene expression and growth rate. *World J. Microbiol. Biotechnol.* 9, 414–420. doi: 10.1007/BF00328029
- Alexander, D. M., and St. John, A. C. (1994). Characterization of the carbon starvation-inducible and stationary phase-inducible gene *slp* encoding an outer membrane lipoprotein in *Escherichia coli*. *Mol. Microbiol.* 11, 1059–1071. doi: 10.1111/j.1365-2958.1994.tb00383.x
- Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001). Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation *in vivo* and *in vitro*. *J. Mol. Biol.* 305, 673–688. doi: 10.1006/jmbi.2000.4327
- Becker, G., and Hengge-Aronis, R. (2004). What makes an *Escherichia coli* promoter  $\sigma^S$  dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of  $\sigma^S$ . *Mol. Microbiol.* 39, 1153–1165. doi: 10.1111/j.1365-2958.2001.02313.x
- Bohannon, D. E., Connell, N., Keener, J., Tormo, A., Espinosa-Urgel, M., Zambrano, M. M., et al. (1991). Stationary-phase-inducible "gearbox" promoters: differential effects of *katF* mutations and role of sigma 70. *J. Bacteriol.* 173, 4482–4492. doi: 10.1128/jb.173.14.4482-4492.1991
- Bruno, J. C. Jr., and Freitag, N. E. (2011). *Listeria monocytogenes* adapts to long-term stationary phase survival without compromising bacterial virulence. *FEMS Microbiol. Lett.* 323, 171–179. doi: 10.1111/j.1574-6968.2011.02373.x

The ugly aspect of this is that such a mechanism results in the persistence of pathogenic bacteria which can cause relapsing of infections. However, the good side is represented by the various biotechnological applications that have come up recently based on the promoters of the genes which are upregulated at stationary phase. In the present review, we have discussed not only the changes at the cellular and molecular levels at stationary phase, but also the various promoters characterized, their regulation and the gene expression systems developed. There are still many unknowns. For example, very little is known about the proteins which are involved in chromosome organization and their interaction with DNA at stationary phase. Such proteins could be important players in regulating gene expression at stationary phase. Further very few SPPs have been experimentally characterized till date. Such promoters should be highly useful for protein production as the growth and protein production phase can be uncoupled. This will pave way toward constructing improved gene expression systems for recombinant protein production.

## AUTHOR CONTRIBUTIONS

JJ and PS wrote and edited the manuscript.

## ACKNOWLEDGMENTS

The authors would like to thank Department of Biotechnology, Government of India for the financial support.

- Busuioac, M., Buttar, B. A., and Piggot, P. J. (2010). The *pdh* operon is expressed in a subpopulation of stationary-phase bacteria and is important for survival of sugar-starved *Streptococcus mutans*. *J. Bacteriol.* 192, 4395–4402. doi: 10.1128/JB.00574-10
- Cao, Y., and Xian, M. (2011). Production of phloroglucinol by *Escherichia coli* using a stationary-phase promoter. *Biotechnol. Lett.* 33, 1853–1858. doi: 10.1007/s10529-011-0638-0
- Cavanagh, A. T., Chandransu, P., and Wassarman, K. M. (2010). 6S RNA regulation of *relA* alters ppGpp levels in early stationary phase. *Microbiology* 156(Pt 12), 3791–3800. doi: 10.1099/mic.0.043992-43990
- Chang, D. E., Smalley, D. J., and Conway, T. (2002). Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol. Microbiol.* 45, 289–306. doi: 10.1046/j.1365-2958.2002.03001.x
- Chang, Y. Y., Wang, A. Y., and Cronan, J. E. Jr. (1994). Expression of *Escherichia coli* pyruvate oxidase (PoxB) depends on the sigma factor encoded by the *rpoS(katF)* gene. *Mol. Microbiol.* 11, 1019–1028. doi: 10.1111/j.1365-2958.1994.tb00380.x
- Chen, H., and Chen, C. Y. (2014). Starvation induces phenotypic diversification and convergent evolution in *Vibrio vulnificus*. *PLOS ONE* 9:e88658. doi: 10.1371/journal.pone.0088658
- Connell, N., Han, Z., Moreno, F., and Kolter, R. (1987). An *E. coli* promoter induced by the cessation of growth. *Mol. Microbiol.* 1, 195–201. doi: 10.1111/j.1365-2958.1987.tb00512.x
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Croxatto, A., Pride, J., Hardman, A., Williams, P., Camara, M., and Milton, D. L. (2004). A distinctive dual-channel quorum-sensing system operates in *Vibrio*



- anguillarum*. *Mol. Microbiol.* 52, 1677–1689. doi: 10.1111/j.1365-2958.2004.04083.x
- Denkin, S. M., and Nelson, D. R. (2004). Regulation of *Vibrio anguillarum* emPA metalloprotease expression and its role in virulence. *Appl. Environ. Microbiol.* 70, 4193–4204. doi: 10.1128/AEM.70.7.4193-4204.2004
- Dersch, P., Schmidt, K., and Bremer, E. (1993). Synthesis of the *Escherichia coli* K-12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. *Mol. Microbiol.* 8, 875–889. doi: 10.1111/j.1365-2958.1993.tb01634.x
- Deutscher, M. P. (2003). Degradation of stable RNA in bacteria. *J. Biol. Chem.* 278, 45041–45044. doi: 10.1074/jbc.R300031200
- Dorman, C. J. (2011). Regulation of transcription by DNA supercoiling in *Mycoplasma genitalium*: global control in the smallest known self-replicating genome. *Mol. Microbiol.* 81, 302–304. doi: 10.1111/j.1365-2958.2011.07718.x
- Espinosa-Urgel, M., Chamizo, C., and Tormo, A. (1996). A consensus structure for sigma S-dependent promoters. *Mol. Microbiol.* 21, 657–659. doi: 10.1111/j.1365-2958.1996.tb02573.x
- Ferrer-Mirallès, N., and Villaverde, A. (2013). Bacterial cell factories for recombinant protein production; expanding the catalogue. *Microb. Cell Fact.* 12:113. doi: 10.1186/1475-2859-12-113
- Finkel, S. E., Zinser, E., Gupta, S., and Kolter, R. (1997). “Life and Death in stationary phase,” in *Molecular Microbiology. NATO ASI Series (Series H: Cell Biology)*, Vol. 103, eds S. J. W. Busby, C. M. Thomas, and N. L. Brown (Berlin: Springer).
- Gaida, S. M., Al-Hinai, M. A., Indurthi, D. C., Nicolaou, S. A., and Papoutsakis, E. T. (2013). Synthetic tolerance: three noncoding small RNAs, DsrA, ArcZ and RprA, acting supra-additively against acid stress. *Nucleic Acids Res.* 41, 8726–8737. doi: 10.1093/nar/gkt651
- Gefen, O., Fridman, O., Ronin, I., and Balaban, N. Q. (2014). Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity. *Proc. Natl. Acad. Sci. U.S.A.* 111, 556–561. doi: 10.1073/pnas.1314114111
- Gosse, J. L., Engel, B. J., Hui, J. C. H., Harwood, C. S., and Flickinger, M. C. (2010). Progress toward a biomimetic leaf: 4,000 h of hydrogen production by coating-stabilized nongrowing photosynthetic *Rhodospseudomonas palustris*. *Biotechnol. Prog.* 26, 907–918. doi: 10.1002/btpr.406
- Gruber, T. M., and Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* 57, 441–466. doi: 10.1146/annurev.micro.57.030502.090913
- Guan, C., Cui, W., Cheng, J., Zhou, L., Guo, J., Hu, X., et al. (2015). Construction and development of an auto-regulatory gene expression system in *Bacillus subtilis*. *Microb. Cell Fact.* 14, 150. doi: 10.1186/s12934-015-0341-2
- Guan, C., Cui, W., Cheng, J., Zhou, L., Liu, Z., and Zhou, Z. (2016). Development of an efficient autoinducible expression system by promoter engineering in *Bacillus subtilis*. *Microb. Cell Fact.* 15, 66. doi: 10.1186/s12934-016-0464-0
- Haldenwang, W. G. (1995). The sigma factors of *Bacillus subtilis*. *Microbiol. Rev.* 59, 1–30.
- Hampshire, T., Soneji, S., Bacon, J., James, B. W., Hinds, J., Laing, K., et al. (2004). Stationary phase gene expression of *Mycobacterium tuberculosis* following a progressive nutrient depletion: a model for persistent organisms? *Tuberculosis* 84, 228–238. doi: 10.1016/j.tube.2003.12.010
- Harms, A., Maisonneuve, E., and Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* 354:aa4268. doi: 10.1126/science.aaf4268
- Hengge, R. (2011). Stationary-phase gene regulation in *Escherichia coli*. *EcoSal Plus* 4. doi: 10.1128/ecosalplus.5.6.3
- Hengge-Aronis, R. (1993). Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in *E. coli*. *Cell* 72, 165–168. doi: 10.1016/0092-8674(93)90655-A
- Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* 66, 373–395. doi: 10.1128/MMBR.66.3.373-395.2002
- Ishihama, A. (1997). Adaptation of gene expression in stationary phase bacteria. *Curr. Opin. Genet. Dev.* 7, 582–588. doi: 10.1016/S0959-437X(97)80003-2
- Jaishankar, J., Singh, P., and Srivastava, P. (2017). Draft genome sequence of a biodesulfurizing bacterium, *Gordonia* sp. strain IITR100. *Genome Announc.* 5:e00230-17. doi: 10.1128/genomeA.00230-17
- Jin, Y., Watt, R. M., Danchin, A., and Huang, J. D. (2009). Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics* 10:165. doi: 10.1186/1471-2164-10-165
- Kim, E. S., Song, J. Y., Kim, D. W., Chater, K. F., and Lee, K. J. (2008). A possible extended family of regulators of sigma factor activity in *Streptomyces coelicolor*. *J. Bacteriol.* 190, 7559–7566. doi: 10.1128/JB.00470-08
- Kim, M. J., Yim, S. S., Choi, J. W., and Jeong, K. J. (2016). Development of a potential stationary-phase specific gene expression system by engineering of SigB-dependent cg3141 promoter in *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 100, 4473–4483. doi: 10.1007/s00253-016-7297-y
- Kolb, A., Kotlarz, D., Kusano, S., and Ishihama, A. (1995). Selectivity of the *Escherichia coli* RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res.* 23, 819–826. doi: 10.1093/nar/23.5.819
- Kolter, R., Siegel, D. A., and Tormo, A. (1993). The stationary phase of the bacterial life cycle. *Annu. Rev. Microbiol.* 47, 855–874. doi: 10.1146/annurev.mi.47.100193.004231
- Lacour, S., and Landini, P. (2004). SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and identification of their promoter sequences. *J. Bacteriol.* 186, 7186–7195. doi: 10.1128/JB.186.21.7186-7195.2004
- Lange, R., Barth, M., and Hengge-Aronis, R. (1993). Complex transcriptional control of the sigma s-dependent stationary-phase-induced and osmotically regulated *osmY* (*csi-5*) gene suggests novel roles for Lrp, cyclic AMP (cAMP) receptor protein-cAMP complex, and integration host factor in the stationary-phase response of *Escherichia coli*. *J. Bacteriol.* 175, 7910–7917. doi: 10.1128/jb.175.24.7910-7917.1993
- Lange, R., and Hengge-Aronis, R. (1994). The *nlpD* gene is located in an operon with *rpoS* on the *Escherichia coli* chromosome and encodes a novel lipoprotein with a potential function in cell wall formation. *Mol. Microbiol.* 13, 733–743. doi: 10.1111/j.1365-2958.1994.tb00466.x
- Lee, S. J., Pan, J. G., Park, S. H., and Choi, S. K. (2010). Development of a stationary phase-specific autoinducible expression system in *Bacillus subtilis*. *J. Biotechnol.* 149, 16–20. doi: 10.1016/j.jbiotec.2010.06.021
- Lemonnier, M., Levin, B. R., Romeo, T., Garner, K., Baquero, M. R., Mercante, J., et al. (2008). The evolution of contact-dependent inhibition in non-growing populations of *Escherichia coli*. *Proc. Biol. Sci.* 275, 3–10. doi: 10.1098/rspb.2007.1234
- Luidalepp, H., Berger, S., Joss, O., Tenson, T., and Polacek, N. (2016). Ribosome shut-down by 16S rRNA fragmentation in stationary-phase *Escherichia coli*. *J. Mol. Biol.* 428(10 Pt B), 2237–2247. doi: 10.1016/j.jmb.2016.01.033
- Marschall, C., Labrousse, V., Kreimer, M., Weichart, D., Kolb, A., and Hengge-Aronis, R. (1998). Molecular analysis of the regulation of *csiD*, a carbon starvation-inducible gene in *Escherichia coli* that is exclusively dependent on sigma s and requires activation by cAMP-CRP. *J. Mol. Biol.* 276, 339–353. doi: 10.1006/jmbi.1997.1533
- Matin, A. (1992). Physiology, molecular biology and applications of the bacterial starvation response. *Soc. Appl. Bacteriol. Symp. Ser.* 21, 49S–57S. doi: 10.1111/j.1365-2672.1992.tb03624.x
- McKay, S. L., and Portnoy, D. A. (2015). Ribosome hibernation facilitates tolerance of stationary-phase bacteria to aminoglycosides. *Antimicrob. Agents Chemother.* 59, 6992–6999. doi: 10.1128/AAC.01532-15
- Melnicki, M., Bianchi, L., Dephillis, R., and Melis, A. (2008). Hydrogen production during stationary phase in purple photosynthetic bacteria. *Int. J. Hydrogen Energy* 33, 6525–6534. doi: 10.1016/j.ijhydene.2008.08.041
- Mengin-Lecreulx, D., and van Heijenoort, J. (1985). Effect of growth conditions on peptidoglycan content and cytoplasmic steps of its biosynthesis in *Escherichia coli*. *J. Bacteriol.* 163, 208–212.
- Navarro Llorens, J. M., Tormo, A., and Martinez-Garcia, E. (2010). Stationary phase in gram-negative bacteria. *FEMS Microbiol. Rev.* 34, 476–495. doi: 10.1111/j.1574-6976.2010.00213.x
- Nystrom, T. (2004). Stationary-phase physiology. *Annu. Rev. Microbiol.* 58, 161–181. doi: 10.1146/annurev.micro.58.030603.123818
- Oliver, J. D. (2005). The viable but nonculturable state in bacteria. *J. Microbiol.* 43, 93–100.
- Ou, J., Wang, L., Ding, X., Du, J., Zhang, Y., Chen, H., et al. (2004). Stationary phase protein overproduction is a fundamental capability of *Escherichia*

- coli*. *Biochem. Biophys. Res. Commun.* 314, 174–180. doi: 10.1016/j.bbrc.2003.12.077
- Pátek, M., and Nešvera, J. (2013). “Promoters and plasmid vectors of *Corynebacterium glutamicum*,” in *Corynebacterium glutamicum. Microbiology Monographs*, Vol. 23, eds H. Yukawa and M. Inui (Berlin: Springer), 51–88.
- Piir, K., Paier, A., Liiv, A., Tenson, T., and Maivali, U. (2011). Ribosome degradation in growing bacteria. *EMBO Rep.* 12, 458–462. doi: 10.1038/embor.2011.47
- Pletnev, P., Osterman, I., Sergiev, P., Bogdanov, A., and Dontsova, O. (2015). Survival guide: *Escherichia coli* in the stationary phase. *Acta Nat.* 7, 22–33.
- Potvin, E., Sanschagrin, F., and Levesque, R. C. (2008). Sigma factors in *Pseudomonas aeruginosa*. *FEMS Microbiol. Rev.* 32, 38–55. doi: 10.1111/j.1574-6976.2007.00092.x
- Qi, Y., Kobayashi, Y., and Hulett, F. M. (1997). The *pst* operon of *Bacillus subtilis* has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the *pho* regulon. *J. Bacteriol.* 179, 2534–2539. doi: 10.1128/jb.179.8.2534-2539.1997
- Ratnayake-Lecamwasam, M., Serron, P., Wong, K. W., and Sonenshein, A. L. (2001). *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev.* 15, 1093–1103. doi: 10.1101/gad.874201
- Rava, P. S., Somma, L., and Steinman, H. M. (1999). Identification of a regulator that controls stationary-phase expression of catalase-peroxidase in *Caulobacter crescentus*. *J. Bacteriol.* 181, 6152–6159.
- Repoila, F., and Gutierrez, C. (1991). Osmotic induction of the periplasmic trehalase in *Escherichia coli* K12: characterization of the *treA* gene promoter. *Mol. Microbiol.* 5, 747–755. doi: 10.1111/j.1365-2958.1991.tb00745.x
- Shaikh, A. S., Tang, Y. J., Mukhopadhyay, A., Martin, H. G., Gin, J., Benke, P. I., et al. (2010). Study of stationary phase metabolism via isotopomer analysis of amino acids from an isolated protein. *Biotechnol. Prog.* 26, 52–56. doi: 10.1002/btpr.325
- Shimada, T., Makinoshima, H., Ogawa, Y., Miki, T., Maeda, M., and Ishihama, A. (2004). Classification and strength measurement of stationary-phase promoters by use of a newly developed promoter cloning vector. *J. Bacteriol.* 186, 7112–7122. doi: 10.1128/JB.186.21.7112-7122.2004
- Singh, P., Chachan, S., Singhi, D., and Srivastava, P. (2016). Isolation and molecular characterization of a stationary phase promoter useful for gene expression in *Gordonia*. *Gene* 591, 153–160. doi: 10.1016/j.gene.2016.07.018
- Su, X., Sun, F., Wang, Y., Hashmi, M. Z., Guo, L., Ding, L., et al. (2015). Identification, characterization and molecular analysis of the viable but nonculturable *Rhodococcus biphenylivorans*. *Sci. Rep.* 5:18590. doi: 10.1038/srep18590
- Tanaka, K., Handel, K., Loewen, P. C., and Takahashi, H. (1997). Identification and analysis of the *rpoS*-dependent promoter of *katE*, encoding catalase HPII in *Escherichia coli*. *Biochim. Biophys. Acta* 1352, 161–166. doi: 10.1016/S0167-4781(97)00044-4
- Tanaka, K., Kusano, S., Fujita, N., Ishihama, A., and Takahashi, H. (1995). Promoter determinants for *Escherichia coli* RNA polymerase holoenzyme containing sigma 38 (the *rpoS* gene product). *Nucleic Acids Res.* 23, 827–834. doi: 10.1093/nar/23.5.827
- Testerman, T. L., Vazquez-Torres, A., Xu, Y., Jones-Carson, J., Libby, S. J., and Fang, F. C. (2002). The alternative sigma factor sigmaE controls antioxidant defences required for *Salmonella* virulence and stationary-phase survival. *Mol. Microbiol.* 43, 771–782. doi: 10.1046/j.1365-2958.2002.02787.x
- Tkachenko, A. G., Kashevarova, N. M., Tyuleneva, E. A., and Shumkov, M. S. (2017). Stationary-phase genes upregulated by polyamines are responsible for the formation of *Escherichia coli* persister cells tolerant to netilmicin. *FEMS Microbiol. Lett.* 364:fnx084. doi: 10.1093/femsle/fnx084
- Tripathi, L., Zhang, Y., and Lin, Z. (2014). Bacterial sigma factors as targets for engineered or synthetic transcriptional control. *Front. Bioeng. Biotechnol.* 2:33. doi: 10.3389/fbioe.2014.00033
- Tunmer, J. R., Robertson, C. R., Schippa, S., and Matin, A. (1992). Use of glucose starvation to limit growth and induce protein production in *Escherichia coli*. *Biotechnol. Bioeng.* 40, 271–279. doi: 10.1002/bit.260400211
- Utsumi, R., Kusafuka, S., Nakayama, T., Tanaka, K., Takayanagi, Y., Takahashi, H., et al. (1993). Stationary phase-specific expression of the *fic* gene in *Escherichia coli* K-12 is controlled by the *rpoS* gene product (sigma 38). *FEMS Microbiol. Lett.* 113, 273–278. doi: 10.1016/0378-1097(93)90217-P
- van de Bunt, B., Bron, P. A., Sijtsma, L., de Vos, W. M., and Hugenholtz, J. (2014). Use of non-growing *Lactococcus lactis* cell suspensions for production of volatile metabolites with direct relevance for flavour formation during dairy fermentations. *Microb. Cell Fact.* 13:176. doi: 10.1186/s12934-014-0176-2
- Wada, A. (1998). Growth phase coupled modulation of *Escherichia coli* ribosomes. *Genes Cells* 3, 203–208. doi: 10.1046/j.1365-2443.1998.00187.x
- Wang, W., Li, X., Wang, J., Xiang, S., Feng, X., and Yang, K. (2013). An engineered strong promoter for streptomycetes. *Appl. Environ. Microbiol.* 79, 4484–4492. doi: 10.1128/AEM.00985-13
- Waterman, S. R., and Small, P. L. (2003). Identification of the promoter regions and sigma(s)-dependent regulation of the *gadA* and *gadBC* genes associated with glutamate-dependent acid resistance in *Shigella flexneri*. *FEMS Microbiol. Lett.* 225, 155–160. doi: 10.1016/S0378-1097(03)00508-1
- Weber, H., Polen, T., Heuveling, J., Wendisch, V. F., and Hengge, R. (2005). Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J. Bacteriol.* 187, 1591–1603. doi: 10.1128/JB.187.5.1591-1603.2005
- Weiner, L., and Model, P. (1994). Role of an *Escherichia coli* stress-response operon in stationary-phase survival. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2191–2195. doi: 10.1073/pnas.91.6.2191
- Wise, A., Brems, R., Ramakrishnan, V., and Villarejo, M. (1996). Sequences in the -35 region of *Escherichia coli* *rpoS*-dependent genes promote transcription by E sigma S. *J. Bacteriol.* 178, 2785–2793. doi: 10.1128/jb.178.10.2785-2793.1996
- Xu, J., and Johnson, R. C. (1995). *aldB*, an *RpoS*-dependent gene in *Escherichia coli* encoding an aldehyde dehydrogenase that is repressed by Fis and activated by Crp. *J. Bacteriol.* 177, 3166–3175. doi: 10.1128/jb.177.11.3166-3175.1995
- Yu, X., Xu, J., Liu, X., Chu, X., Wang, P., Tian, J., et al. (2015). Identification of a highly efficient stationary phase promoter in *Bacillus subtilis*. *Sci. Rep.* 5:18405. doi: 10.1038/srep18405
- Zambrano, M. M., and Kolter, R. (1996). GASping for life in stationary phase. *Cell* 86, 181–184. doi: 10.1016/S0092-8674(00)80089-6

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Jaishankar and Srivastava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Features of CRISPR-Cas Regulation Key to Highly Efficient and Temporally-Specific crRNA Production

Andjela Rodic<sup>1,2</sup>, Bojana Blagojevic<sup>3</sup>, Magdalena Djordjevic<sup>3</sup>, Konstantin Severinov<sup>4,5</sup> and Marko Djordjevic<sup>1\*</sup>

<sup>1</sup> Faculty of Biology, Institute of Physiology and Biochemistry, University of Belgrade, Belgrade, Serbia, <sup>2</sup> Multidisciplinary PhD Program in Biophysics, University of Belgrade, Belgrade, Serbia, <sup>3</sup> Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia, <sup>4</sup> Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, United States, <sup>5</sup> Skolkovo Institute of Science and Technology, Skolkovo, Russia

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
United States

### Reviewed by:

Jintao Liu,  
University of California, San Diego,  
United States  
Robert Martin Blumenthal,  
University of Toledo, United States  
Andrea Ciliberto,  
IFOM - The FIRC Institute of Molecular  
Oncology, Italy

### \*Correspondence:

Marko Djordjevic  
dmarko@bio.bg.ac.rs

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 11 July 2017

**Accepted:** 19 October 2017

**Published:** 03 November 2017

### Citation:

Rodic A, Blagojevic B, Djordjevic M,  
Severinov K and Djordjevic M (2017)  
Features of CRISPR-Cas Regulation  
Key to Highly Efficient and  
Temporally-Specific crRNA  
Production. *Front. Microbiol.* 8:2139.  
doi: 10.3389/fmicb.2017.02139

Bacterial immune systems, such as CRISPR-Cas or restriction-modification (R-M) systems, affect bacterial pathogenicity and antibiotic resistance by modulating horizontal gene flow. A model system for CRISPR-Cas regulation, the Type I-E system from *Escherichia coli*, is silent under standard laboratory conditions and experimentally observing the dynamics of CRISPR-Cas activation is challenging. Two characteristic features of CRISPR-Cas regulation in *E. coli* are cooperative transcription repression of cas gene and CRISPR array promoters, and fast non-specific degradation of full length CRISPR transcripts (pre-crRNA). In this work, we use computational modeling to understand how these features affect the system expression dynamics. Signaling which leads to CRISPR-Cas activation is currently unknown, so to bypass this step, we here propose a conceptual setup for cas expression activation, where cas genes are put under transcription control typical for a restriction-modification (R-M) system and then introduced into a cell. Known transcription regulation of an R-M system is used as a proxy for currently unknown CRISPR-Cas transcription control, as both systems are characterized by high cooperativity, which is likely related to similar dynamical constraints of their function. We find that the two characteristic CRISPR-Cas control features are responsible for its temporally-specific dynamical response, so that the system makes a steep (switch-like) transition from OFF to ON state with a time-delay controlled by pre-crRNA degradation rate. We furthermore find that cooperative transcription regulation qualitatively leads to a cross-over to a regime where, at higher pre-crRNA processing rates, crRNA generation approaches the limit of an infinitely abrupt system induction. We propose that these dynamical properties are associated with rapid expression of CRISPR-Cas components and efficient protection of bacterial cells against foreign DNA. In terms of synthetic applications, the setup proposed here should allow highly efficient expression of small RNAs in a narrow time interval, with a specified time-delay with respect to the signal onset.

**Keywords:** CRISPR-Cas activation, pre-crRNA processing, CRISPR regulation, crRNA generation, biophysical modeling

## INTRODUCTION

CRISPR-Cas are adaptive immune systems, which defend prokaryotic cells against foreign DNA, including viruses and plasmids. A CRISPR-Cas system consists of a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) array and associated *cas* genes (Makarova et al., 2006; Barrangou et al., 2007; Brouns et al., 2008; Hille and Charpentier, 2016). CRISPR arrays consist of identical direct repeats (R) of about 30 bp in length, interspaced with spacers (S) of similar length and variable sequence. Spacer sequences are often complementary to fragments of viral or plasmid DNA. A match between a CRISPR spacer and invading phage (bacterial virus) sequence provides immunity to infection (Barrangou et al., 2007; Hille and Charpentier, 2016). The entire CRISPR locus is initially transcribed as a long transcript (called pre-crRNA) (Pougach et al., 2010; Pul et al., 2010), which is further processed by Cas proteins to small protective CRISPR RNAs (called crRNAs) (Brouns et al., 2008; Pougach et al., 2010; Djordjevic et al., 2012). crRNAs are responsible for recognition and, together with Cas proteins, inactivation of invading foreign genetic elements (Brouns et al., 2008; Al-Attar et al., 2011). Cas proteins also take part in CRISPR adaptation, which is a process in which new spacers from viral genomes are inserted in CRISPR array. **Figure 1** shows a schematic gene diagram for Type I-E CRISPR-Cas from *E. coli*, (Mojica and Diez-Villasenor, 2010; Patterson et al., 2017), which we consider in this paper. The *cas* genes and the CRISPR array are transcribed from separate promoters, which are located inside of the intergenic regions here denoted by IGLB and L (the leader sequence), respectively (see **Figure 1**; Pougach et al., 2010; Pul et al., 2010).

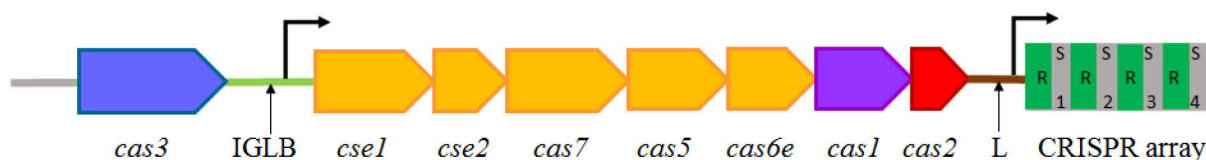
Promoters for *cas* operon and the CRISPR array are repressed in Type I-E CRISPR-Cas in *E. coli* (Pougach et al., 2010; Pul et al., 2010; Westra et al., 2010), which makes this system silent under standard conditions. Consequently, to generate crRNAs that can protect the bacterial cell, CRISPR-Cas has to be activated. Thus, to understand the system function it is crucial to understand the main features that control dynamics of CRISPR-Cas activation (Mojica and Diez-Villasenor, 2010; Richter et al., 2012; Patterson et al., 2017). However, approaching this problem experimentally is complicated due to the following:

- i. It requires direct experimental observation of *in vivo* dynamics of molecular species (proteins or RNA) in a cell (see e.g., Morozova et al., 2015).
- ii. The signaling which leads to system induction is currently unclear (Ratner et al., 2015; Patterson et al., 2017), e.g., even a viral infection, an obvious trigger, is not sufficient to activate the system.
- iii. To understand the roles of the key system features in its response/dynamics these features would have to be perturbed, which may require extensive reengineering of the system.

A complementary approach is to use mathematical/biophysical modeling to assess how different features of CRISPR-Cas expression affect system dynamics. Moreover, *in silico* analysis allows one to study alternative system architectures, and/or to perturb the natural system (see e.g., Rodic et al., 2017), which in turn allows understanding the role of its key regulatory features.

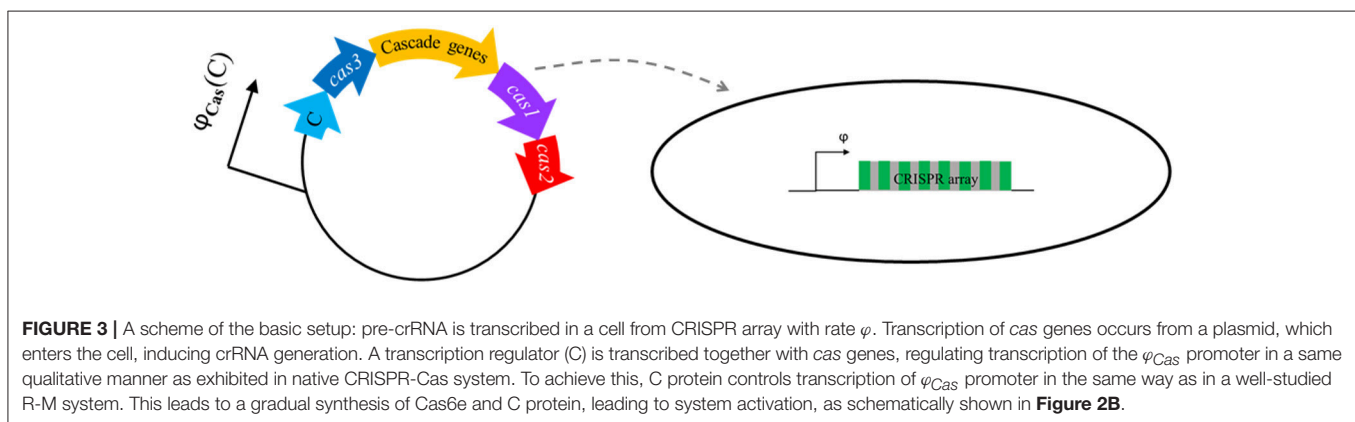
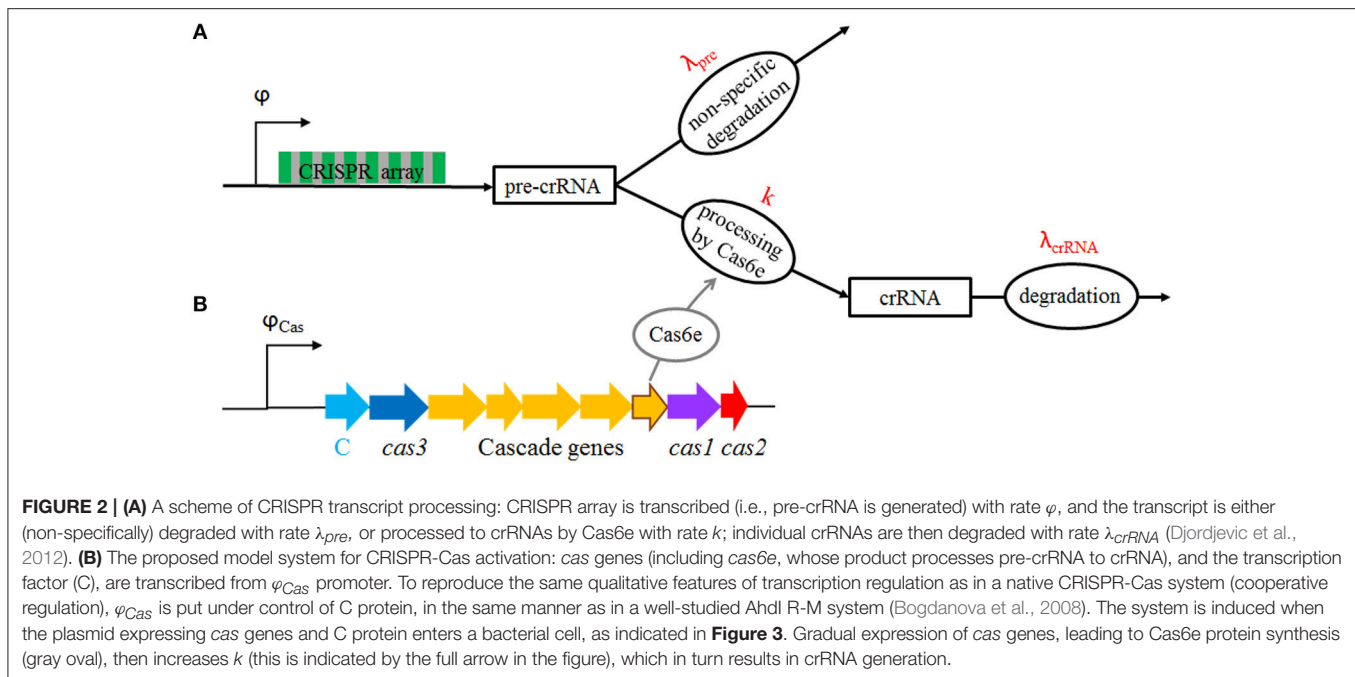
Experimental research has led to a consistent picture of the main CRISPR-Cas regulatory features in closely related *E. coli* and *Salmonella enterica* (Pul et al., 2010; Westra et al., 2010; Medina-Aparicio et al., 2011). Under standard conditions, promoters for both CRISPR array and *cas* genes are repressed by global regulators (H-NS and LRP). Repression by these regulators is highly cooperative, as their binding is nucleated at certain position, and then extends along the DNA through cooperative interactions between repressor molecules (Bouffartigues et al., 2007). Additional regulators, such as CRP, may also be involved in the repression of *cas* operon (Yang et al., 2014). While the exact signaling mechanism remains unclear, this repression must be relieved upon appropriate external signal (e.g., envelope stress that may signal bacteriophage invasion), through the action of transcription activators (LexA, LeuO, and BaeR-S are likely involved) (Richter et al., 2012; Patterson et al., 2017). In particular, for Type I-E CRISPR-Cas in *E. coli*, it was shown that cooperative repression by H-NS can be relieved by elevated amount of LeuO (Pul et al., 2010; Westra et al., 2010). Thus, highly cooperative repression, which is abolished by transcription activators, emerges as a major feature of CRISPR-Cas transcription control in *E. coli* and its relatives.

Another crucial mechanism in CRISPR-Cas expression is pre-crRNA transcript processing (Brouns et al., 2008; Pougach et al., 2010). Experiments in *E. coli*, reported that overexpression of



**FIGURE 1** | A scheme of a Type I-E CRISPR-Cas system from *E. coli* (Al-Attar et al., 2011; Makarova et al., 2011, 2015). The *cas* genes and the CRISPR array are indicated. R and S within the CRISPR array correspond, respectively, to repeats and spacers; note that the spacer sequences differ from each other, and are labeled by consecutive numbers (1, 2, 3...). IGLB and L correspond to intergenic regions where promoters for the Cascade complex genes (*cse1,2*, *cas7,5,6e*) and the Cas1,2 adaptation proteins (IGLB) and the CRISPR array (L) are located. The two promoters within IGLB and L are indicated by arrows. One of the Cas proteins (Cas6e) is responsible for processing pre-crRNA to crRNA. The effector Cascade complex is composed of proteins encoded by genes marked with yellow color. It binds crRNA, which recognizes invading DNA. Once recognized, foreign DNA is destroyed by the product of *cas3* (Brouns et al., 2008).





Cas6e (which is responsible for pre-crRNA processing) generates highly abundant crRNAs from pre-crRNA which is present at low abundance (Pougach et al., 2010). We previously showed that a simple quantitative model—whose relevant kinetic scheme is shown in **Figure 2A**—explains this observation (Djordjevic et al., 2012), so that a small decrease in pre-crRNA abundance leads to a much larger (around two orders of magnitude) increase in crRNA abundance. Interestingly, the main mechanism responsible for this strong amplification is fast non-specific degradation of pre-crRNA (see **Figure 2**) by unidentified nuclease(s). In particular, when *cas* genes expression increases, processing of pre-crRNA by Cas6e is favored and diverts the entire pre-crRNA molecule away from the path of non-specific degradation. Therefore, the fast non-specific degradation of pre-crRNA should be considered as a second major regulatory feature of CRISPR-Cas expression.

The modeling described in Djordjevic et al. (2012) took into account only the transcript processing step, i.e., it was assumed that there is an infinitely abrupt (stepwise) increase of pre-crRNA

to crRNA processing rate, and pre-crRNA generation rate. This is, however, a clear idealization of the induction mechanism, as transcription regulation of *cas* genes and CRISPR array promoters is neglected. That is, in reality, pre-crRNA processing rate can be increased only gradually, as it takes time to synthesize the needed Cas proteins. The rate of Cas proteins synthesis is in turn directly related to the transcription control of the *cas* gene promoter in the IGLB region (see **Figure 1**). Similarly, the rate by which pre-crRNA is synthesized is determined by the transcription control of the CRISPR array promoter (L region).

Consequently, a more realistic model of CRISPR-Cas expression dynamics has to take into account both the regulation of CRISPR array and Cas protein synthesis, and CRISPR transcript processing. However, a major obstacle in achieving such model is that signaling which leads to the system induction, and detailed mechanism of CRISPR-Cas transcription regulation, is still unclear. We here propose a model system for CRISPR-Cas induction by assuming that activation of crRNA production

is put under transcriptional control exhibited in a restriction-modification (R-M) immune system (Pingoud et al., 2014). As argued below, such model system would have qualitative features of transcription regulation expected for a CRISPR-Cas, and will keep the same transcript processing mechanism as that described for native system. On the other hand, this model system allows bypassing the currently unknown signaling that leads to CRISPR-Cas activation, and can be readily analyzed *in silico*, since transcription regulation of a well-studied R-M system (AhdI, see Bogdanova et al., 2008)—for which we previously showed that it can be reliably modeled (see below)—is used as a proxy for transcription regulation of CRISPR-Cas system.

Through this approach, we expect to:

- i. Obtain quantitatively more realistic model of CRISPR-Cas induction dynamics, in which the transcription regulation, i.e., the gradual synthesis of relevant enzymes and transcription regulators is explicitly taken into account.
- ii. Qualitatively understand the main features of CRISPR-Cas induction, in particular the roles of cooperative transcription regulation, and of fast non-specific degradation of pre-crRNA.
- iii. Propose an experimental setup for CRISPR-Cas induction that mimics the main qualitative features of the native system.

The setup of the model will be explicitly considered in the next subsection.

## RESULTS

### *In silico* Experiment Setup

#### The Model System

We start from a CRISPR transcript processing scheme, which is shown in **Figure 2**. According to this scheme, pre-crRNA is generated with rate  $\varphi$ , and subsequently either non-specifically degraded (due to activity of an unspecified nuclease) with rate  $\lambda_{pre}$ , or is processed by Cas6e to crRNAs with rate  $k$ . crRNAs are subsequently degraded with rate  $\lambda_{crRNA}$ . All the parameters in the scheme are experimentally determined in (Djordjevic et al., 2012) (for Type I-E CRISPR-Cas in *E. coli*) and explicitly stated in Methods. In particular, the main feature of the transcript processing is a large (non-specific) pre-crRNA degradation rate (with  $\lambda_{pre} \sim 1$  1/min), which is much larger than crRNA degradation rate (with  $\lambda_{crRNA} \sim 1/100$  1/min). In the experiments, crRNA production is artificially activated, by overexpressing Cas6e from a plasmid, which increases pre-crRNA processing rate ( $k$ ) for between one and two orders of magnitude (between  $10\lambda_{pre}$  and  $100\lambda_{pre}$ ). While the repression of the *cas* promoter in IGLB region (see **Figure 1**) is very strong, with very small amount of Cas6e synthesized when the system is uninduced, the repression of the CRISPR array promoter is significantly weaker, with rather strong basal rate of pre-crRNA generation ( $\varphi \sim 10$  1/min) (Pougach et al., 2010; Pul et al., 2010; Westra et al., 2010; Djordjevic et al., 2012).

As indicated in the Introduction, we previously modeled the transcript processing mechanism (Djordjevic et al., 2012), where we took that  $k$  is increased abruptly, i.e., as a step function at  $t = 0$ . This neglects the transcription regulation of *cas*

and CRISPR array promoters. Such abrupt increase of  $k$  will provide a baseline for our predictions, which will now take into account that Cas6e (the enzyme which processes pre-crRNA to crRNA) is synthesized gradually. While in the experiments crRNA generation is activated by overexpressing Cas6e from a plasmid (see e.g., Pougach et al., 2010), it is likely that in the native system the expression of CRISPR array is activated as well (Pul et al., 2010). Consequently, we will also take into account a gradual synthesis of the regulator [in our case, a C-protein (Tao et al., 1991; Bogdanova et al., 2008)], which can activate CRISPR array transcription by increasing the basal rate  $\varphi$  to a higher value.

To include transcription regulation of the *cas* promoter, i.e., the gradual synthesis of Cas6e and C transcriptional regulator, we here propose the model system whose setup is schematically shown in **Figures 2, 3**. This setup includes a CRISPR array which is expressed from a promoter with basal transcription activity  $\varphi$  (**Figure 3**). The second component is a vector (plasmid, virus) which expresses *cas* genes and the control protein C that are jointly transcribed from a promoter with transcription activity  $\varphi_{Cas}$ . While Cas3 is not directly relevant for the problem considered here (dynamics of crRNA generation), as it does not take part in crRNA biogenesis, it is necessary for CRISPR interference (Hille and Charpentier, 2016). We therefore include it in the setup to allow expression of all *cas* genes, i.e. to have a fully functional CRISPR-Cas system.

As detailed below,  $\varphi_{Cas}$  is regulated by C. To mimic the qualitative features of transcription regulation in native CRISPR-Cas system, we employ the transcription regulation found in some R-M systems, as explained in the next subsection. The system is activated when the vector enters a bacterial cell lacking its own *cas* genes, which leads to a gradual synthesis of Cas proteins (including Cas6e), therefore increasing the processing rate  $k$ , which in turn leads to crRNA generation (see **Figure 2B**—the full arrow) by pre-crRNA processing. Gradual increase of pre-crRNA generation rate can be also considered through this model, through activation of CRISPR array promoter by gradually synthesized C.

Note that the setup above, where *cas* genes are introduced in a cell on a vector, allows bypassing the unknown signaling step in CRISPR-Cas induction. That is, the vector entering the cell marks the start of the system activation (setting zero time in the dynamics simulations), and mimics the signaling which starts synthesis of the transcription activator. Therefore, the key regulatory features which characterize the downstream steps (CRISPR array transcription and transcript processing) can be studied both *in silico* (which will be done here), and also potentially experimentally. In terms of experimental implementation, introducing *cas* genes in a cell on a virus also allows synchronizing the cell population, which is an approach previously implemented to visualize R-M protein kinetics (Mruk and Blumenthal, 2008).

### Putting CRISPR-Cas under Transcription Control of an R-M System

As discussed above, *cas* promoter will be put under transcription control exhibited by R-M systems. Below, the main elements

necessary for modeling the system transcription regulation are introduced.

R-M systems are often mobile, and can spread from one bacterial host to the other (Mruk and Kobayashi, 2013). When a plasmid carrying R-M system genes enters a naive bacterial host, the host genome is initially unmethylated, and can consequently be cut by the restriction enzyme. It is, therefore, evident that expression of the restriction enzyme and methyltransferase must be tightly regulated in order to ensure that bacterial genome is protected by the methyltransferase (“antidote”), before it is cut by the restriction enzyme. This tight regulation is often achieved through a dedicated control (C) proteins (Tao et al., 1991; Viješuric et al., 2000).

We here concentrate on the AhdI R-M system, whose transcription control by C protein has been well-studied (Bogdanova et al., 2008). The activation of AhdI by C protein is reminiscent of CRISPR-Cas activation, as strong cooperative interactions are involved in both cases. In particular, C proteins bound at promoter-proximal and promoter-distal operators interact with high binding cooperativity, so that configuration in which only one operator is occupied cannot be observed in the absence of RNA polymerase (RNAP). At lower C protein concentrations, RNAP can outcompete C protein bound at promoter-proximal operator, leading to transcriptionally active configuration (Bogdanova et al., 2009). Moreover, another feature exhibited in AhdI transcription control, i.e., autoregulation by C protein, is also likely found in CRISPR-Cas transcription regulation. That is, *LeuO* that activates CRISPR-Cas expression (Westra et al., 2010) also regulates its own transcription. In particular, similarly to transcription regulation of *cas* genes, *leuO* is repressed by H-NS, while this repression is abolished by *LeuO* (Chen et al., 2001). At high concentrations, C protein is bound at both promoter-proximal and promoter-distal position, leading to the promoter repression—see **Figure 5** in (Bogdanova et al., 2009) and the scheme of the transcription configurations shown in **Figure 5** (framed in the figure). Negative autoregulation is also exhibited by *LeuO*, as it inhibits transcription activation of its gene by BglJ-RcsB (Stratmann et al., 2012). Therefore, putting *cas* genes under transcription control found in AhdI mimics the main qualitative features of CRISPR-Cas transcription regulation, namely, gradual synthesis of Cas proteins, cooperativity in transcription regulation, and putative autoregulation.

Another advantage of this setup is that we previously showed that biophysical modeling can be used to: (i) explain *in vitro* measurements of the *wild type* and mutant R-M system transcription control (Bogdanova et al., 2008), (ii) explain *in vivo* measurements of the system dynamics (Morozova et al., 2015), (iii) effectively perturb the main R-M system features and relate these perturbations with the system dynamics (Rodic et al., 2017). Consequently, transcription control of a well-studied AhdI R-M system, whose transcription regulation can be reliably modeled (Bogdanova et al., 2008), will serve as a proxy for the transcription control of a much less understood CRISPR-Cas system.

## In silico Analysis of the Main System Features

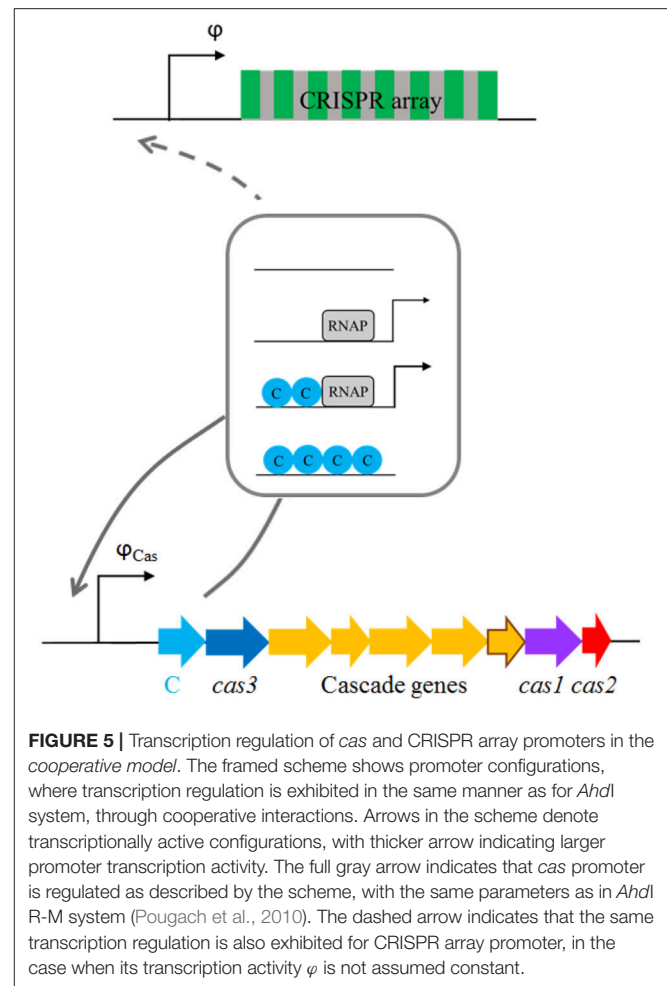
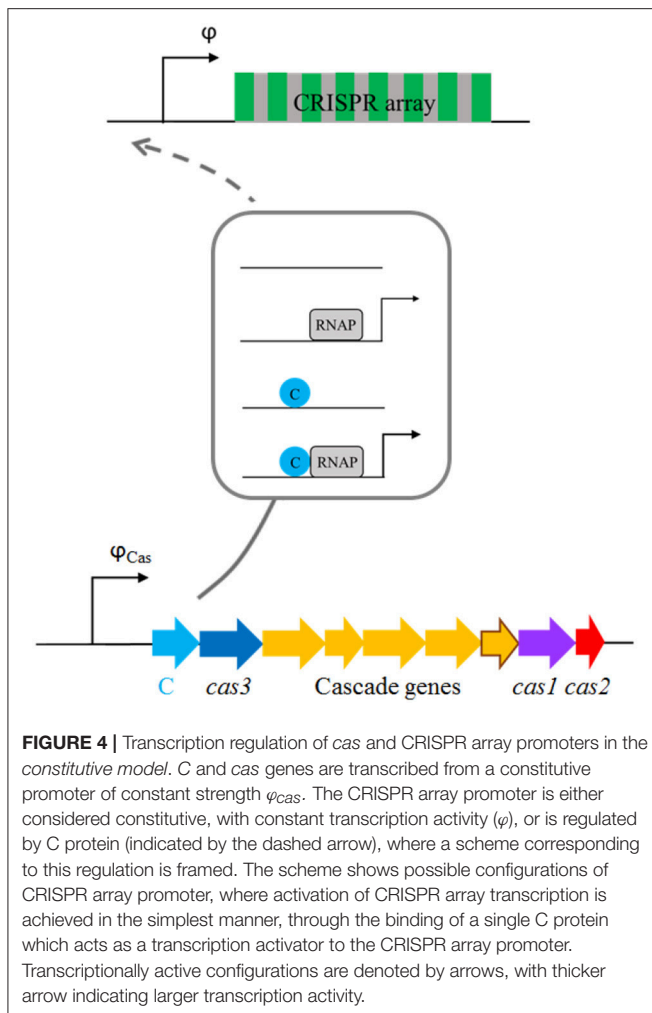
The baseline for our predictions will be provided by a model in which the increase of pre-crRNA to crRNA processing rate  $k$  is infinitely abrupt—we will call this the baseline model. Comparing the baseline model with predictions that take into account the system transcription regulation (as schematically shown in **Figures 2, 3**), allows analyzing how gradual synthesis of Cas6e affects kinetics of crRNA generation.

While in the native CRISPR-Cas both *cas* genes and CRISPR array promoters are repressed by global regulators, the repression of *cas* genes was found to be much stronger (Pul et al., 2010; Westra et al., 2010)—consequently, when the system is (experimentally) artificially induced, this is commonly done by expressing only *cas* genes (Pougach et al., 2010; Semenova et al., 2016; Musharova et al., 2017). However, in the native system, it is likely that expression of both CRISPR array and *cas* genes is activated when the appropriate induction signal(s) is received (Pul et al., 2010). We will therefore investigate the system dynamics when only *cas* genes are activated (i.e., only pre-crRNA processing rate is gradually increased), and when *cas* genes and CRISPR array promoter transcription are jointly (and gradually) increased. Consequently, in both of the models introduced below (constitutive and cooperative), we will consider two options. First, when only transcription of *cas* genes is activated, while transcription activity of CRISPR array remains constant. Second, we will consider the case when the transcription activity of CRISPR array is increased as well.

We further introduce two models of *cas* gene and CRISPR array transcription regulation:

- i *The constitutive model* (**Figure 4**). In this model *cas* genes are expressed from a constitutive promoter, so that they are transcribed with the constant rate once the plasmid enters a cell. In the case when we consider that the system is activated by only increasing pre-crRNA processing rate, the transcription activity  $\varphi$  is kept constant. When CRISPR array transcription rate is increased as well, increasing  $\varphi$  is exhibited in the simplest manner, by binding of a single C protein activator. Note that, in accordance with its name, no cooperativity is exhibited for transcription regulation described by this model.
- ii *The cooperative model* (**Figure 5**). In this model, C protein regulates the transcription of *cas* genes, and its own transcription, in the same manner as in AhdI R-M system. As noted above, such transcription regulation is characterized by strong cooperative interactions. CRISPR array transcription rate is either kept constant, or in the case when it is increased, we take that it is exhibited in the same way as for *cas* promoter transcription (the dashed arrow in **Figure 5**).

Studying of the two models allows one to assess how the cooperative transcription regulation (which also characterizes the native CRISPR-Cas system) compares to the activation in which no cooperativity is exhibited, and therefore allows us to assess the role of this key system feature. Also, considering the two models when  $\varphi$  is first kept constant, and then increased together with  $k$ , allows assessing significance of CRISPR array transcription



control. To allow a direct comparison of models dynamics, the overall strength of  $\varphi_{Cas}$  is adjusted so that the same value of maximal pre-crRNA processing rate is achieved. Similarly, when the transcription rate of CRISPR array is increased, the interaction parameters are adjusted so that the same equilibrium increase of  $\varphi$  is achieved in both models (see Methods).

## Modeling Results

### Kinetics of Pre-crRNA and crRNA Production

We first consider the situation in which crRNA generation is activated by expressing Cas proteins, such that the processing rate  $k$  is gradually increased, while the CRISPR array transcription activity remains constant. In this case, we compare the system dynamics for: (i) *baseline model*, in which the processing rate  $k$  is increased as a step function, which corresponds to the limit of infinitely fast system induction, (ii) *constitutive model* (see **Figure 4**), and (iii) *cooperative model* (see **Figure 5**).

In constitutive and cooperative models, the gradual synthesis of Cas6e leads to gradual change of transcript processing rate  $k$  ( $k^*$  is a processing constant):

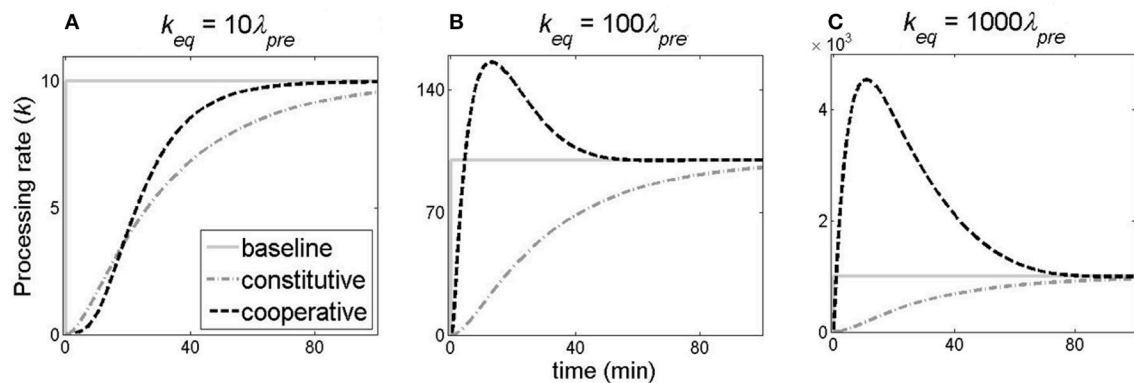
$$k(t) = [Cas6e](t) \cdot k^* \quad (1)$$

**Figure 6** illustrates how the processing rate ( $k$ ) changes with time, when the baseline, constitutive, and cooperative models of *cas* gene expression are assumed. For the constitutive model (the dash-dotted curve), the processing rate uniformly increases and reaches an equilibrium value, for all values of  $k_{eq}$  considered in three panels of **Figure 6**. On the other hand, for cooperative model (the dashed curve) and at higher values of  $k_{eq}$  (**Figures 6B,C**), we see a rapid increase of  $k$  at initial times, followed by a fast return to the equilibrium value due to repression at higher *C* protein concentrations.

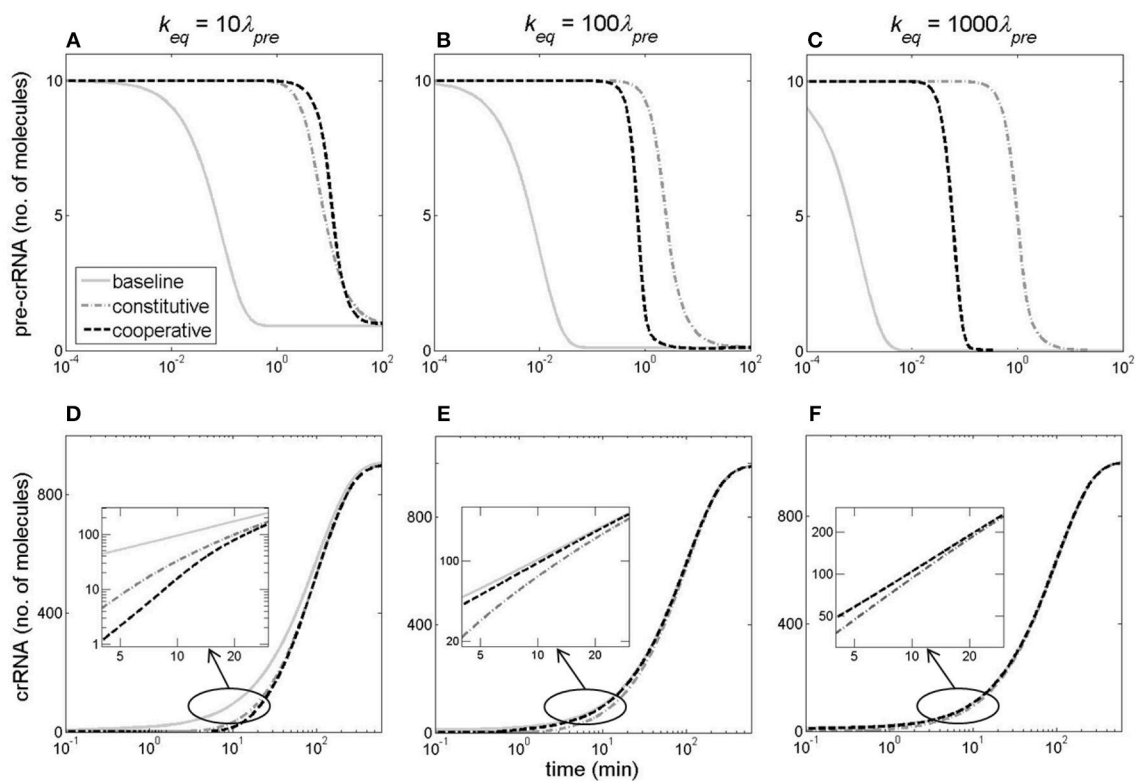
In **Figure 7**, we address how different  $k$  dynamics (shown in **Figure 6**), affects pre-crRNA and crRNA generation. Specifically,  $\varphi$  is held constant at its initial value (10 1/min), while  $k$  changes according to the baseline, constitutive, or cooperative models until reaching the same equilibrium value of  $10\lambda_{pre}$ ,  $100\lambda_{pre}$ , and  $1,000\lambda_{pre}$  (left, central, and right columns of **Figure 7**, respectively). The model of abrupt Cas6e expression serves as a baseline for assessing the dynamics in the other two models (constitutive and cooperative), in which Cas6e is realistically (gradually) expressed.

In **Figures 7A–D**, we see that cooperative model leads to the steepest transition from ON to OFF state (in the case





**FIGURE 6** | Comparing the dynamics of the pre-crRNA processing rate change. The change of the processing rate  $k$  with time is shown for: the baseline model (solid gray curve), the constitutive model (dash-dotted gray curve) and the cooperative model (dashed black curve). (A–C) correspond to different  $k_{eq}$  values ( $k_{eq} = 10\lambda_{pre}$ ,  $100\lambda_{pre}$ ,  $1,000\lambda_{pre}$ , respectively). CRISPR transcription activity is constant ( $10\ 1/\text{min}$ ).



**FIGURE 7** | Kinetics of pre-crRNA and crRNA generation. The columns correspond to  $k_{eq}$  values of  $10\lambda_{pre}$  (A,D),  $100\lambda_{pre}$  (B,E), and  $1,000\lambda_{pre}$  (C,F), which are reached through the baseline model (the gray solid curve), the constitutive model (the gray dash-dotted curve) or cooperative model (the black dashed curve). The upper (A–C) and the lower (D–F) rows correspond, respectively, to pre-crRNA and crRNA kinetics. CRISPR array promoter transcription activity is kept constant at  $10\ 1/\text{min}$ .

of pre-crRNA), and from OFF to ON state (in the case of crRNA). Furthermore, we can distinguish between two different regimes in **Figure 7**. At lower  $k_{eq}$  (left column in **Figure 7**), there is a noticeably slower accumulation of crRNA at early times in both cooperative and constitutive models compared to the

baseline model of infinitely abrupt processing rate ( $k$ ) increase (**Figure 7D**). On the other hand, at higher  $k_{eq}$  ( $k_{eq} \geq 100\ 1/\text{min}$ , the central and right columns in **Figure 7**), the dynamics of crRNA accumulation for cooperative model becomes faster compared to constitutive model dynamics at early times, and

approaches the limit of infinitely abrupt  $k$  increase (see the inserts in **Figures 7E,F**). The faster kinetics of crRNA increase in cooperative model is due to the fast increase of  $k$  at early times in this model (**Figures 6B,C**).

### Effects of *cas* Genes Regulation

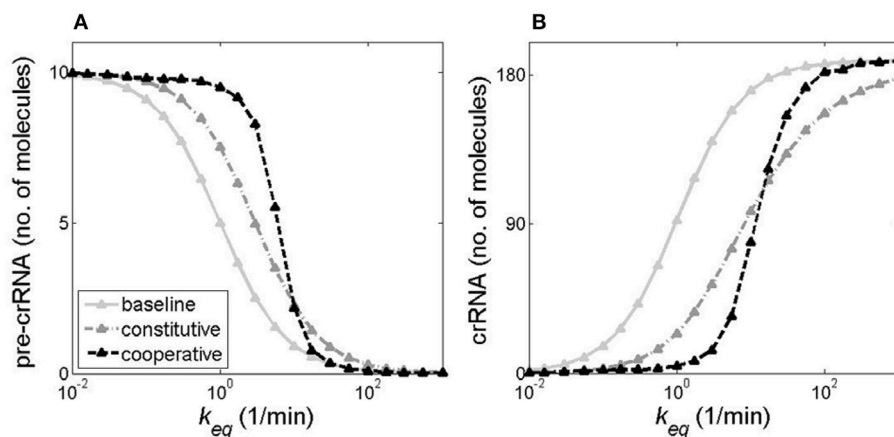
From **Figure 7**, we observe that transcripts reach their steady-state levels quite late, i.e.,  $>100$  min post-induction. Such late time is, however, not relevant for cell response to phage infection, since infected *E. coli* lyse  $\sim 20$  min post-infection, while shut-off of essential cell functions happens earlier (Kruger and Schroeder, 1981). Therefore, in **Figure 8** we estimate pre-crRNA and crRNA levels for all three models at 20 min post-induction, as the maximal value of pre-crRNA processing rate  $k_{eq}$  is changed from very low to high values ( $>100\lambda_{pre}$ , characteristic for artificial Cas6e induction), while keeping the level of CRISPR array transcription constant ( $\varphi = 10$  1/min).

The following features emerge from **Figure 8**:

- i. A switch-like system behavior for both pre-crRNA and crRNA curves in the cooperative model, while the constitutive and baseline models yield much more gradual responses to changes in  $k_{eq}$ . For crRNA, the cooperative model leads to a rapid transition from the OFF state (with essentially no crRNA generated at 20 min), to the ON state (with high abundance of crRNA), and reciprocal situation for pre-crRNA. Consequently, for small amounts of synthesized Cas6e (i.e., small  $k_{eq}$  values), which can be caused by leaks in *cas* promoter activity, the system remains in OFF state. On the other hand, once the system is activated when the processing rate (directly related to the amount of Cas6e available) reaches a certain threshold ( $k_{eq} \gtrsim 50$ ), a large amount of crRNA is generated at early times, which should allow protection from foreign DNA invasion. The significance of this behavior is considered in Discussion.
- ii. An interesting cross-over behavior in the cooperative model, where at low  $k_{eq}$  values crRNA amounts are low, while at high  $k_{eq}$  values the synthesized crRNA amounts become larger than in the constitutive model, and approach the baseline model curve. Therefore, at high  $k$ -values ( $\sim 100$  1/min), which are encountered in experiments, (Pougach et al., 2010; Djordjevic et al., 2012) the model of cooperative *cas* gene expression leads to accumulation of protective crRNA amounts close to those achievable in the limit of infinitely abrupt  $k$  increase. Consequently, the high cooperativity in transcription regulation, characteristic for native CRISPR-Cas system regulation, leads to a highly efficient crRNA generation at high transcript processing rates.
- iii. Sufficient crRNA levels are generated to protect host cell against bacteriophage infection, at early times post-induction, even at relatively low values of pre-crRNA processing rate. That is,  $k_{eq}$  somewhat larger than 11/min leads to  $\sim 10$  crRNAs which already corresponds to the amount that negatively affects phage development (Pougach et al., 2010); moreover, a small additional  $k_{eq}$  increase leads to a large increase in generated crRNAs in the cooperative model, due to the rapid transition from OFF to ON state.
- iv. A saturation in generated crRNA amounts at early times post-induction. That is, for  $k_{eq} \sim 100$  1/min the amount of generated crRNAs at 20 min stops significantly increasing with further increase in  $k_{eq}$ . This saturation can be relieved (leading to increase in the amount of generated crRNA), if CRISPR array transcription activity is increased, which is further analyzed below.

### Perturbing Pre-crRNA Degradation Rate

We next perturb the second key feature of CRISPR-Cas regulation—fast non-specific degradation of pre-crRNA. The consequence of pre-crRNA degradation rate  $\lambda_{pre}$  decrease at constant  $\varphi$  was next investigated for all three models. The



**FIGURE 8 |** Pre-crRNA and crRNA amounts early post-induction for different models of *cas* gene transcription regulation. The figure shows **(A)** pre-crRNA and **(B)** crRNA amounts 20 min post-induction (i.e., 20 min after introduction of the vector expressing *cas* genes), as a function of the maximal (equilibrium) value of the transcript processing rate  $k$ . CRISPR promoter transcription activity is kept constant ( $\varphi = 10$  1/min). The gray solid, the gray dash-dotted, and the black dashed curves correspond, respectively, to baseline, constitutive, and cooperative models of *cas* regulation.

decrease was followed at different  $k_{eq}$  values (i.e., at different levels of Cas6e activity), where  $\varphi$  is held constant.

The effects of  $\lambda_{pre}$  decrease are similar for all three models, so in **Figure 9** we show the results only for the cooperative model. For all  $k_{eq}$  values we see that abolishing the fast decay of pre-crRNA (decreasing  $\lambda_{pre}$ ), significantly decreases the time delay of the onset of crRNA generation. This effect is most pronounced at high  $k_{eq}$  values (**Figure 9C**). Also, perturbing the degradation rate deforms crRNA dynamics curve with respect to the standard Hill (sigmoidal) shape that is exhibited at high  $\lambda_{pre}$  such as  $\lambda_{pre} = 1/50$ . Furthermore, analogously to **Figure 8**, in **Figure S1** (Supplementary Material), we show how crRNA amount at 20 min after induction depends on pre-crRNA degradation rate  $\lambda_{pre}$ . One can clearly observe that as  $\lambda_{pre}$  decreases, the amount of generated crRNA early post-induction significantly increases, consistently with the decrease of the time delay of onset of crRNA generation observed in **Figure 9**.

### Relieving crRNA Production Saturation by Increasing Pre-crRNA Generation

In addition to *cas* genes, CRISPR array promoter is also repressed (though more weakly) by global transcription regulators (Pul et al., 2010; Westra et al., 2010). Consequently, crRNA generation can be also augmented by increasing CRISPR array transcription activity. Therefore, we next assess how joint increase of  $k$  (achieved by activating *cas* gene transcription) and  $\varphi$  (achieved by increasing CRISPR array transcription) affects generated crRNA amount 20 min post-induction for all three regulatory models.

As can be seen from **Figure 10**, increasing  $\varphi$  robustly relieves crRNA saturation (see also discussion of **Figure 8**). Moreover, one can see that a relatively modest, factor of two increase of  $\varphi$  (from 10 1/min to 20 1/min) can abolish the need of a significant, order of magnitude,  $k$  increase to produce the same amount or crRNA. As above, we observe a switch-like behavior for the cooperative model (compare **Figure 10C** with **Figures 10A,B**),

with cooperative model curves exhibiting the steepest transition from OFF to ON state for all  $\varphi$  values.

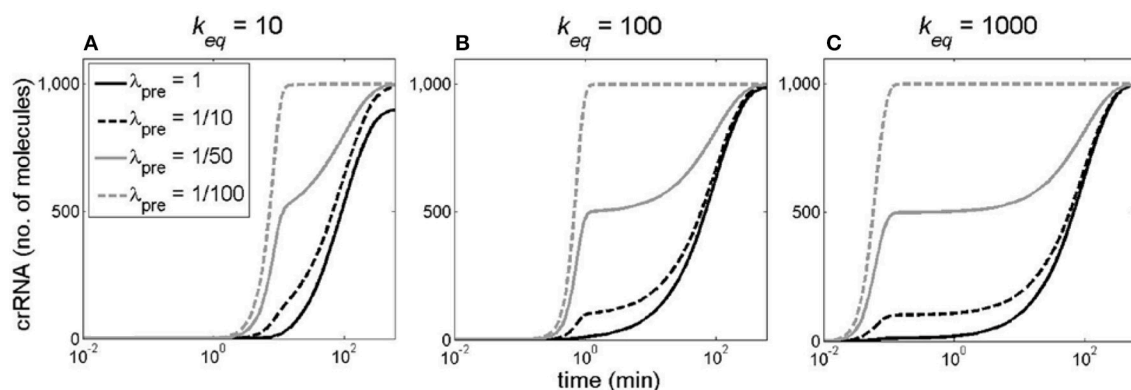
### Regulation of CRISPR Array Transcription Activity

We next consider how different models of regulation of CRISPR array transcription affect crRNA dynamics. For all three models, the transcription activity  $\varphi$  is increased by an order of magnitude (from  $\varphi = 10$  1/min to  $\varphi = 100$  1/min), for different  $k_{eq}$  values ( $k_{eq} = \lambda_{pre}$ ,  $10\lambda_{pre}$ , and  $100\lambda_{pre}$ ), see **Figure S2** (Supplementary Material). We obtain that the cooperative model leads to a more controlled (attenuated) pre-crRNA dynamics, which is due to the presence of repressing mechanism at high C protein amounts (see **Figure S3**). For crRNA dynamics, we observe that the cooperative model exhibits the steepest transition from OFF to ON state. Moreover, this model leads to the largest delay in crRNA generation. Consequently, in addition to pre-crRNA degradation rate, the cooperative transcription regulation also contributes to the delay between the activating signal and the onset of crRNA generation.

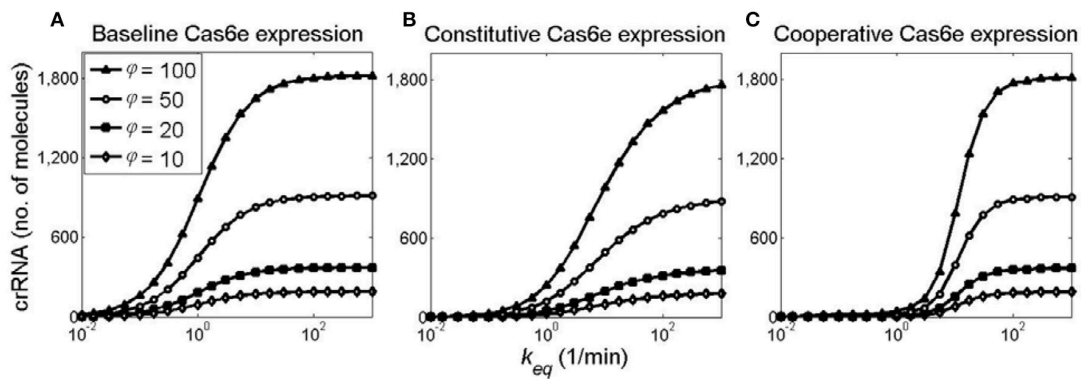
We previously (**Figure 9**) perturbed pre-crRNA degradation rate while keeping the transcription rate  $\varphi$  constant. Finally, we now also decrease  $\lambda_{pre}$  under the conditions when both *cas* genes and CRISPR array transcription is activated according to all three models (see **Figure S4**). The results are qualitatively similar to **Figure 9** (where  $\varphi$  is constant), i.e., decreasing  $\lambda_{pre}$  diminishes the switch-like system response and/or decreases the time-delay in the onset of pre-crRNA generation.

## DISCUSSION AND SUMMARY

One of the most prominent problems in understanding CRISPR-Cas function is assessing dynamics of the system activation, i.e., understanding the roles of the key features of CRISPR-Cas regulation. Addressing this problem is complicated by the fact that exact conditions for system activation remain unclear. In fact, for Type I-E CRISPR-Cas system in *E. coli*, even bacteriophage infection itself is not sufficient



**FIGURE 9 |** The effect of perturbing pre-crRNA degradation rate on the dynamics of crRNA generation. The pre-crRNA processing rate increases to its equilibrium value through the cooperative model, while  $\varphi$  is held constant (at 10 1/min). Different curves correspond to different  $\lambda_{pre}$  values: 1 (solid black), 1/10 (dashed black), 1/50 (solid gray), and 1/100 1/min (dashed gray). (**A–C**), correspond to different  $k_{eq}$  values indicated at the top of each panel.



**FIGURE 10 |** Relieving saturation in generated crRNA amounts through joint  $k$  and  $\varphi$  increase. crRNA amount as a function of  $k_{eq}$  at 20 min post-induction, obtained for (A) baseline, (B) constitutive, and (C) cooperative models. Curves marked with diamonds, filled squares, circles, and filled triangles, correspond, respectively, to  $\varphi$  of 10, 20, 50, and 100 1/min.

to induce the system. We here proposed a synthetic setup which allows inducing CRISPR-Cas with qualitative features that correspond to native system regulation, while bypassing currently unclear conditions under which the system is activated. This setup involves putting *cas* genes and/or CRISPR array under transcription control found in a well-studied R-M system, which exhibits cooperative transcription regulation that is also characteristic of CRISPR-Cas regulation (Bouffartigues et al., 2007; Westra et al., 2010). A major advantage of the setup is that it can be readily experimentally implemented, e.g., by introducing *cas* genes and the regulator (C protein) in a cell on a virus. This would allow synchronizing the cell population, and experimentally observing the system dynamics, where such measurements could be directly compared with the predictions provided here. Another advantage is that major parameters in the setup have been inferred from experimental data, as both CRISPR transcript processing, and AhdI transcription regulation, have been experimentally well-studied (Bogdanova et al., 2008; Pougach et al., 2010; Djordjevic et al., 2012).

Consequently, this setup allows us to directly (*in silico*) address how the system regulation contributes to its dynamical response. In particular, previous experimental and computational work point to cooperative regulation of *cas* gene and CRISPR array transcription, and fast non-specific degradation of pre-crRNA, as two main system regulatory features (Pougach et al., 2010; Pul et al., 2010; Westra et al., 2010; Djordjevic et al., 2012). We therefore investigated two alternative regulatory architectures, one with constitutive, and the other with cooperative *cas* gene regulation. The dynamics corresponding to these two architectures was then compared with the baseline model, in which pre-crRNA processing rate is increased infinitely abruptly. We assessed the dynamics in the case when only *cas* genes are activated (i.e., only pre-crRNA processing rate is gradually increased), and when *cas* genes and CRISPR array promoter transcription is jointly increased. We focused on early system dynamics (within the first 20 min post-induction), as this period is most relevant for defending the cell against invading viruses. Finally, we also perturbed the high pre-crRNA non-specific

degradation rate, under different system conditions described above, and assessed what effect such perturbation has on system dynamics.

The main result of the analysis is that the system regulation leads to a clear switch-like behavior, characterized by an initial delay of crRNA synthesis, followed by a steep transition from OFF to ON state. Unexpectedly, it is not only the cooperative transcription regulation, but also fast non-specific pre-crRNA degradation, which leads to such dynamics. That is, decreasing the high pre-crRNA degradation rate effectively abolishes the delay in crRNA generation, and deforms the crRNA kinetics from the standard sigmoidal (Hill) shape (Hill, 2013) typical for switch-like system response (Figure 9). Interestingly, we also found that, when pre-crRNA processing rate and CRISPR array transcription rate are jointly (and gradually) increased, as likely exhibited in the native system, the system is more robust to perturbations in the degradation rate (Figure S4).

The cooperative transcription regulation leads to an interesting cross-over behavior in the early system dynamics. At low pre-crRNA processing rates, cooperative regulation leads to much smaller crRNA amounts at early times compared to constitutive expression. On the other hand, at higher processing rates, there is a large increase in synthesized crRNA amounts, which approach the limit of infinitely abrupt system induction. Interestingly, when the system is artificially activated by overexpressing *cas* genes, pre-crRNA processing rates correspond to the regime of the highly enhanced crRNA production (Djordjevic et al., 2012). While the parameters of the native system induction are unclear, it is tempting to hypothesize that they may also reach this cross-over, allowing the system to generate crRNAs with the rate close to the limit of infinitely fast induction at times when they are needed.

The rapid transition of the system from OFF to ON state is straightforward to interpret in terms of its function in immune response. When a potential signal indicating infection is received by the cell, CRISPR-Cas has a very short time to generate sufficient crRNA amounts to protect the cell, as bacteriophages are typically highly efficient in shutting-down essential cell functions. Thus, there is a question whether enough crRNA can



be generated in a model which accounts for gradual synthesis of proteins that process pre-crRNA and/or are responsible for gradual CRISPR array activation. We robustly obtained that enough crRNA can be generated at early times, even when the system is activated by only increasing the pre-crRNA processing rate. Moreover, a much smaller increase of the processing rate is needed to achieve certain crRNA amount, if CRISPR array transcription is activated as well. Therefore, these results may explain the relatively inefficient repression of CRISPR array promoter, since even a small increase of CRISPR array transcription rate efficiently increases generated crRNA amounts. In fact, the need to rapidly produce large amounts of crRNAs may be a major constraint on system dynamics.

In distinction to the rapid transition of the system from “OFF” to “ON” state, interpretation of the delay in crRNA generation, which comes as a model prediction, is less straightforward. One possibility is that such delay is related with primed adaptation in CRISPR-Cas, which relies on a pre-existing (priming) spacer that enables a biased uptake of new spacers—therefore serving to minimize infection by phage escape mutants that would otherwise evade the interference (Sternberg et al., 2016). In particular, it has been found that priming is facilitated by slow or delayed CRISPR interference, leading to a steady-state flux of substrates from which new spacers can be acquired (Kunne et al., 2016; Severinov et al., 2016; Musharova et al., 2017). Such delay in CRISPR interference can clearly be achieved by a delay in crRNA generation that is predicted in our work.

It has been proposed that Type I-E CRISPR-Cas in *E. coli* may have functions other than immunity. For example, it was found by bioinformatics analysis that the system is changing very slowly, in distinction to rapid diversification of CRISPR arrays in other species, indicating that the system is not taking an active role in defense against immediate viral threats (Touchon et al., 2011). In this respect, it may be useful to view the dynamical properties inferred here in a more general terms, namely of a capability of expressing a large number of molecules in a narrow time interval, with a specific time-delay with respect to reception of an external signal. It is clear that such highly efficient, and temporally specific response, may be highly desirable for multiple cellular functions. It would be very interesting to find out how functions of *E. coli* Type I-E CRISPR-Cas, yet to be discovered in the future, would fit within the dynamical properties inferred here.

## METHODS

We start from a previously introduced model of CRISPR transcript processing by Cas proteins (Djordjevic et al., 2012). In this model (see **Figure 2A**), a short-living transcript [pre-crRNA] is synthesized with a promoter transcription activity  $\varphi$ , and further, either quickly degraded with a degradation rate  $\lambda_{pre}$ , or processed (cut) into shorter, long-living RNAs [crRNA] with a processing rate  $k$ . Processed transcripts are degraded with a rate  $\lambda_{crRNA}$ . In the equations below, we assume that the processing rate depends linearly on the substrate (pre-crRNA) amount, since the amount of pre-crRNA is small [ $<10$  molecules per cell (Pougach et al., 2010)], so that the corresponding kinetic

equations are:

$$\frac{d[\text{pre-crRNA}]}{dt} = \varphi - (\lambda_{pre} + k) \cdot [\text{pre-crRNA}] \quad (2)$$

$$\frac{d[\text{crRNA}]}{dt} = k \cdot [\text{pre-crRNA}] - \lambda_{crRNA} \cdot [\text{crRNA}] \quad (3)$$

The equations above are further solved deterministically, as both CRISPR array and *cas* genes are expressed from promoters with strong basal transcription. Furthermore, the small pre-crRNA amount is due to fast non-specific degradation, i.e., due to the transcript processing step. With respect to this, note that there is an excess of enzyme (Cas6e) over substrate (pre-crRNA) (Djordjevic et al., 2012), so the equations describing the transcript processing are linear. Therefore, their deterministic solution accurately describes the mean of the stochastic simulations.

In the previous study (Djordjevic et al., 2012), we considered a model in which transcription regulation is neglected, so that  $k$  and  $\varphi$  increase in an idealized manner, i.e., infinitely abruptly. We now introduce models where the relevant enzymes and transcription regulators are synthesized in a realistic (i.e., gradual) manner. Specifically,  $k$  in Equation now explicitly depends on time, and is proportional to the enzyme (the processing protein, Cas6e) concentration, i.e.,  $k = [\text{Cas6e}] \cdot k^*$ , where  $k^*$  is processing constant. We here consider that this processing rate  $k$  can change with time in the following ways:

- 1) Infinitely abruptly, from 0 to its equilibrium value,  $k_{eq}$  at  $t = 0$ , which we refer to as the baseline model.
- 2) Gradually, with  $[\text{Cas6e}](t)$ , where Cas6e is expressed from a constitutive promoter (promoter with constant transcription activity), see **Figure 4**.
- 3) Also gradually with  $[\text{Cas6e}](t)$ , where Cas6e is now expressed from an AhdI-like regulated promoter (see **Figure 5**).

As noted above, we either keep the CRISPR array transcription rate  $\varphi$  constant (which allows us investigating the dynamics in response to changing only pre-crRNA processing rate), or allow  $\varphi$  to change:

- 1) Infinitely abruptly (the baseline model), so that at  $t = 0$  it increases from its starting value (10 1/min) to the equilibrium value.
- 2) Gradually, through the simplest activation mechanism, where a single C protein activates transcription from the CRISPR array promoter (the dashed arrow in **Figure 4**).
- 3) Also gradually with  $C(t)$ , where the same transcription regulation as in AhdI RM system is exhibited (the dashed arrow in **Figure 5**).

In constructing Cas6e and CRISPR expression models, we refer to our existing model of AhdI restriction-modification (RM) system control (Bogdanova et al., 2008), which describes expression of the control protein (C) and the restriction endonuclease (R)—C and R are co-transcribed in AhdI RM system. We here use a thermodynamical model of CR operon transcription regulation, and a dynamical model of transcript and protein expression.

For  $t = 0$  we take the moment when plasmid carrying C and *cas* genes enters the naïve host. Thus, all initial conditions are set to zero, except for  $[pre-crRNA](t = 0) = \varphi/\lambda_{pre} = 10 \text{ (1/min)}$  (Djordjevic et al., 2012), as extracted from the Equation in equilibrium. Note that while C and *cas* genes enter the cell on a plasmid, CRISPR array is expressed within the cell, with the transcription rate  $\varphi$ .

## Constitutive Model of *cas* Gene and CRISPR Array Expression

We assume that C and *cas* genes are co-transcribed from a constitutive (unregulated) *cas* promoter (see above and Figure 4). C and *cas* transcript and protein concentrations change with time:

$$\frac{d[c - cas](t)}{dt} = \varphi_{Cas} - \lambda_{Cas} \cdot [c - cas](t) \quad (4)$$

$$\frac{dC(t)}{dt} = k_C \cdot [c - cas](t) - \lambda_C \cdot C(t) \quad (5)$$

$$\frac{d[Cas6e](t)}{dt} = k_{Cas6e} \cdot [c - cas](t) - \lambda_{Cas6e} \cdot [Cas6e](t). \quad (6)$$

Note that all the notation (including in the equation above), is introduced in Table 1. The first terms on the right-hand side represent transcript/protein synthesis by transcription/translation, while the second terms represent transcript/protein decay by degradation. The parameter values are as in AhdI RM system model (with Cas6e now replacing R in AhdI system), and are also provided in the table at the end of the methods. Since C and Cas6e protein degradation rates are taken to be the same, it follows:

$$[Cas6e](t) = \frac{k_{Cas6e}}{k_C} C(t), \quad (7)$$

So that the differential equation for Cas6e dynamics can be omitted. We set the value of  $\varphi_{Cas}$  to one (see the next subsection) so that the equilibrium processing rate is the same for the constitutive and the cooperative models (see e.g., Figure 6), which allows a direct comparison of the dynamics in these two models. Consequently, we set  $k^*$  so that  $k_{eq} = [Cas6e]_{eq} \cdot k^* = 10 \text{ (1/min)}$ . Regarding CRISPR array transcription  $\varphi$ , we keep it constant, in the case when we consider the system activation by overexpression of *cas* genes. In the case when we also consider activation of CRISPR transcription, we introduce a simple model of CRISPR expression regulation (the dashed arrow in Figure 4), where CRISPR promoter, apart from being unoccupied, can be found in the following three configurations, which are represented by the reactions shown below: (i) RNAP alone bound to the promoter (8), (ii) a C monomer alone bound to its binding site (9), and (iii) RNAP recruited by a C monomer bound to its binding site, acting as a transcription activator —note that these configurations correspond to the second, third and fourth line in the framed part of Figure 4, respectively.

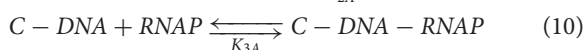
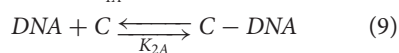
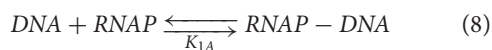


TABLE 1 | Notations used in model equations.

Variables	Description	
$\varphi_{Cas}$	Transcription activity of <i>cas</i> promoter	
$\varphi$	Transcription activity of CRISPR promoter	
$[c - cas]$	Concentration of <i>cas</i> operon transcript	
$[pre-crRNA]$	Concentration of unprocessed CRISPR array transcript	
$[crRNA]$	Concentration of processed CRISPR array transcript	
C	Concentration of control protein	
$[Cas6e]$	Concentration of processing protein	
KINETIC MODEL CONSTANTS		
$k^*$	CRISPR transcript processing constant	0.02
$k_C$	Translation constant for control protein	0.60
$k_{Cas6e}$	Translation constant for processing protein	3.00
$\lambda_{Cas}$	Rate of <i>cas</i> transcript decay	0.20
$\lambda_{pre}$	Rate of unprocessed CRISPR transcript decay	1.00
$\lambda_{crRNA}$	Rate of processed CRISPR transcript decay	0.01
$\lambda_C$	Rate of control protein decay	0.033
$\lambda_{Cas6e}$	Rate of Cas6e processing protein decay	0.033
TRANSCRIPTION REGULATION MODELS CONSTANTS		
$\alpha$	Proportionality constants	1.663
$\gamma$		110
$\alpha'$		110
$a$	Constants which absorb the relevant	$1.60 \times 10^{-1}$
$p$	equilibrium dissociation constants and RNA	$9.25 \times 10^{-1}$
$q$	polymerase concentration	$1.41 \times 10^{-5}$
$d$		$1.00 \times 10^{-1}$
$e$		Adjusted
$f$		$2.00 \times 10^2$
$a'$		$1.00 \times 10^{-1}$
$p'$		Adjusted
$q'$		$2.50 \times 10^{-5}$
$K_D$		$6.50 \times 10^2$

The equilibrium dissociation constants of the above reactions are given by:

$$K_{1A} = [DNA][RNAP] / [RNAP - DNA] \quad (11)$$

$$K_{2A} = [DNA][C] / [C - DNA] \quad (12)$$

$$K_{3A} = [C - DNA][RNAP] / [C - DNA - RNAP]. \quad (13)$$

Using the Shea-Ackers based approach, i.e. assuming that the transcription activity is proportional to the equilibrium promoter occupancy by RNAP, we derive the expression for CRISPR promoter transcriptional activity:

$$\varphi = \gamma \frac{Z_{RNAP} + Z_{C-RNAP}}{1 + Z_{RNAP} + Z_C + Z_{C-RNAP}} \quad (14)$$

where  $\gamma$  is a proportionality constant, while configuration statistical weights correspond to:  $Z_{RNAP} = [RNAP - DNA] / [DNA]$  — RNAP alone bound to the promoter,  $Z_C = [C - DNA] / [DNA]$  — C monomer alone bound to its binding site,  $Z_{C-RNAP} = [C - DNA - RNAP] / [DNA]$  — RNAP

recruited to the promoter by a bound C monomer. We can obtain  $\varphi$  dependence on C concentration:

$$\varphi(C) = \gamma \frac{d + def[C]}{1 + d + e[C] + def[C]} \quad (15)$$

If we introduce parameters expressed in terms of the equilibrium binding constants and RNAP concentration:

$$d = [RNAP]/K_{1A} \quad (16)$$

$$e = 1/K_{2A} \quad (17)$$

$$f = K_{1A}/K_{3A}. \quad (18)$$

To estimate the parameters, we use a condition:

$$\varphi(0) = 10 \frac{1}{\min} \quad (19)$$

which corresponds to the value in Djordjevic et al. (2012), and:

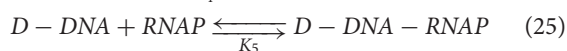
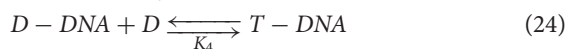
$$\varphi(C_{eq}) = 100 \frac{1}{\min} \quad (20)$$

Another (evident) condition is that the fraction, which appears on the right-hand side of the Equation (15), has to be smaller than 1. By adjusting the parameters to satisfy the conditions (19) and (20), we obtain  $d < 1/9$ , which allows setting the values of  $d$  and  $\gamma$ . Further, we notice that  $e = 99 / ([C]_{eq} \cdot (f - 100))$  and, having fixed the value of  $f$ , we can adjust  $e$  with respect to  $[C]_{eq}$ .

The unprocessed [pre-crRNA] and processed [crRNA] transcript amounts change with time according to the Equations (2) and (3), where  $\varphi$  is given by.

## Cooperative Model of *cas* and CRISPR Expression

As opposed to the constitutive *cas* operon expression, we here assume that the *cas* promoter is regulated by C as in the wild type AhdI RM system (Bogdanova et al., 2008), through cooperative interactions (see **Figure 5**). The following set of reactions describes the transcriptional regulation of the *cas* promoter by the C protein (note the promoter configurations shown in **Figure 5**):



where C and D stand for C protein monomers and dimers, respectively.

The reactions (21)–(25) represent:

- (21) C monomers dimerization;
- (22) RNAP binding to the *cas* promoter forming RNAP-DNA complex;

- (23) D binding to the distal binding site forming D-DNA complex;
- (24) D recruitment to the proximal binding site forming T-DNA complex;
- (25) RNAP recruitment to the *cas* promoter forming D-DNA-RNAP complex.

In equilibrium the above reactions lead to the following equations of the equilibrium dissociation constants:

$$K_1 = \frac{[C]^2}{[D]} \quad (26)$$

$$K_2 = \frac{[DNA][RNAP]}{[RNAP - DNA]} \quad (27)$$

$$K_3 = \frac{[D][DNA]}{[D - DNA]} \quad (28)$$

$$K_4 = \frac{[D][D - DNA]}{[T - DNA]} \quad (29)$$

$$K_5 = \frac{[RNAP][D - DNA]}{[D - DNA - RNAP]} \quad (30)$$

Taking into account the aforementioned Shea-Ackers assumption we obtain:

$$\varphi_{Cas} = \alpha \frac{Z_{RNAP} + Z_{D-RNAP}}{1 + Z_{RNAP} + Z_{D-RNAP} + Z_T}, \quad (31)$$

$\alpha$  is a proportionality constant,  $Z_{RNAP} = [RNAP - DNA]/[DNA]$ ,  $Z_{D-RNAP} = [D - DNA - RNAP]/[DNA]$  and  $Z_T = [T - DNA]/[DNA]$  denote the statistical weights of only RNAP bound to the promoter, RNAP recruited to the promoter by a C dimer bound to the distal binding site, and a C tetramer repressing transcription, respectively.

By using Equations (26)–(30), the Equation (31) can be rewritten in terms of C monomer concentration (following the notation in Bogdanova et al., 2008; Rodic et al., 2017):

$$\varphi_{Cas}(C) = \alpha \frac{a + b[C]^2}{1 + a + b[C]^2 + c[C]^4} \quad (32)$$

which can be expressed, by using the redefined parameters, in the following form:

$$\varphi_{Cas}(C) = \alpha \frac{a + ap[C]^2}{1 + a + ap[C]^2 + p^2q[C]^4}. \quad (33)$$

We set  $\alpha$  so that the equilibrium value of *cas* transcription activity corresponds to one (adapted from Bogdanova et al., 2008). Parameters  $a$ ,  $p$ , and  $q$  depend on the equilibrium dissociation constants and RNAP concentration and are given by:

$$a = [RNAP]/K_2 \quad (34)$$

$$p = \frac{K_2}{K_1 K_3 K_5} \quad (35)$$

$$q = \frac{1}{K_1^2 K_3 K_4 p^2} = \frac{K_3 K_5^2}{K_2^2 K_4} \quad (36)$$

While their values are deduced from the already determined  $a$ ,  $b$ , and  $c$ , that correspond to the best fit to the AhdI experimentally measured transcription activity vs.  $C$  (Bogdanova et al., 2008).

Regarding the dynamics, note that  $C$  and Cas6e transcript and protein amounts change with time according to the Equations (4)–(6), where  $\varphi_{Cas}$  is given by.

Similarly as for the constitutive model, we keep  $\varphi$  constant, in the case when we consider inducing the system through increasing pre-crRNA processing rate. When we also consider regulation of CRISPR array transcription, we assume that CRISPR promoter is regulated by  $C$  in the same way as  $cas$  promoter. Thus, following the same procedure we obtain for the CRISPR promoter transcription activity:

$$\varphi = \alpha' \frac{a' + a'p'[C]^2}{1 + a' + a'p'[C]^2 + p'^2q'[C]^4} \quad (37)$$

where constants  $\alpha'$ ,  $a'$ ,  $p'$ , and  $q'$  are determined by imposing the same constraints on  $\varphi$  as above (-). Specifically, these constraints lead to the condition  $a' < \frac{1}{9}$ , which allows setting parameters  $a'$  and  $\alpha'$ . Further, from Equation (20) we express  $p'$  in terms of  $q'$  and get  $q' < \frac{1}{400 \times 99}$  (deduced from the real roots criterion of quadratic equation), based on which we set  $q'$ , and subsequently obtain the relation for adjusting  $p'$  with respect to  $k_{eq}$  (i.e.,  $C_{eq}$ ). Again, the unprocessed [pre-crRNA] and processed [crRNA] transcript amounts change with time according to the Equations (2) and (3), where  $\varphi$  is replaced with (37).

## Changing Pre-crRNA Processing Rate

From Equation (1) we have that

$$k_{eq} = [Cas6e]_{eq} \cdot k^*, \quad (38)$$

where we adjust the equilibrium value of  $k$  in the constitutive and the cooperative case by varying the concentration of Cas6e in equilibrium. The equilibrium Cas6e concentration can be derived from the steady-state conditions for Equations and :

$$[Cas6e]_{eq} = \frac{k_{Cas6e}}{\lambda_{Cas} \lambda_{Cas6e}} \varphi_{Cas}(C_{eq}). \quad (39)$$

In the model of constitutive  $C$  and Cas6e expression, the equilibrium concentration of Cas6e is adjusted through the change of  $\varphi_{Cas}$  (being constant with time). In the case of cooperative  $C$  and Cas6e expression,  $[Cas6e]_{eq}$  is adjusted through the change of  $\alpha$  in Equation (33), i.e., through the change of overall  $cas$  promoter strength, taking into account that  $[C]_{eq}$  is proportional to  $[Cas6e]_{eq}$  according to (7).

## Joint Change of $k$ and $\varphi$

We here investigate how the joint change of  $k$  and  $\varphi$ , which corresponds to the joint increase of  $cas6e$  and CRISPR array gene expression, affects the dynamics of [pre-crRNA] and [crRNA] transcripts. We start from the baseline model of infinitely abrupt increase of  $k$  and  $\varphi$ . We then compare the baseline model to the more realistic case of constitutive and the cooperative models. We take  $\varphi$  change from the initial value of 10 1/min to 100 1/min in equilibrium, while  $k_{eq}$  takes on values  $\lambda_{pre}$ ,  $10\lambda_{pre}$ , and

$100\lambda_{pre}$ . Note that the change in  $k_{eq}$ , implies joint change of  $\varphi_{Cas}$  in Equation (4) and  $e$  in Equation (15) in the constitutive case; in the cooperative case it implies joint change of  $\alpha$  and  $p$  in Equation (33) and  $p'$  in Equation (37), which ensures the same functional dependency  $\varphi(t)$ , for different values of  $k_{eq}$ .

## Perturbing Pre-crRNA Degradation Rate

### $\lambda_{pre}$

The pre-crRNA degradation rate  $\lambda_{pre}$  is perturbed (decreased) in the following two cases:

- With the transcription rate  $\varphi$  (10 1/min) held constant. The equilibrium value of  $k$  is then adjusted by varying  $\varphi_{Cas}$  in the constitutive, and  $\alpha$  in the cooperative model.
- When both  $\varphi$  and the processing rate  $k$  reach the equilibrium value (100 1/min) gradually, with the effect of the change assessed in all three models (baseline, constitutive and cooperative).  $k_{eq}$  reaches the value 100 1/min through the change of  $\varphi_{Cas}$  in the constitutive, and  $\alpha$  and  $p$  in the cooperative model, while  $\varphi$  increases from  $\varphi(0) = 10$  1/min to  $\varphi(C_{eq}) = 1,001$ /min through adjusting the parameters  $e$  in the constitutive, and  $p'$  in the cooperative model.

Note that changing  $\lambda_{pre}$  affects the initial amount of pre-crRNA (which is an initial condition for the differential equations) according to the relation  $[pre - crRNA]_{eq}(t = 0) = \varphi(t = 0) / \lambda_{pre}$  (see Equation 2), which follows from the steady-state condition for pre-crRNA when the system is not activated.

## AUTHOR CONTRIBUTIONS

All authors have given approval to the final version of the manuscript. MarD conceived and coordinated the work, with the help of KS and MagD. AR and BB performed calculations and the analysis. All the authors interpreted the results and contributed to writing the manuscript.

## FUNDING

This work was funded by the Swiss National Science foundation under SCOPES project number IZ73Z0\_152297, by Marie Curie International Reintegration Grant within the 7th European community Framework Programme (PIRG08-GA-2010-276996) and by the Ministry of Education and Science of the Republic of Serbia under project number ON173052. KS acknowledges support through NIH grant RO1 GM10407.

## ACKNOWLEDGMENTS

We thank Ekaterina Semenova for carefully reading the paper and useful suggestions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02139/full#supplementary-material>



## REFERENCES

- Al-Attar, S., Westra, E. R., van der Oost, J., and Brouns, S. J. (2011). Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* 392, 277–289. doi: 10.1515/bc.2011.042
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140
- Bogdanova, E., Djordjevic, M., Papapanagiotou, I., Heyduk, T., Kneale, G., and Severinov, K. (2008). Transcription regulation of the type II restriction-modification system AhdI. *Nucleic Acids Res.* 36, 1429–1442. doi: 10.1093/nar/gkm1116
- Bogdanova, E., Zakharova, M., Streeter, S., Taylor, J., Heyduk, T., Kneale, G., et al. (2009). Transcription regulation of restriction-modification system Esp1396I. *Nucleic Acids Res.* 37, 3354–3366. doi: 10.1093/nar/gkp210
- Bouffartigues, E., Buckle, M., Badaut, C., Travers, A., and Rimsky, S. (2007). H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.* 14, 441–448. doi: 10.1038/nsmb1233
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964. doi: 10.1126/science.1159689
- Chen, C. C., Fang, M., Majumder, A., and Wu, H. Y. (2001). A 72-base pair AT-rich DNA sequence element functions as a bacterial gene silencer. *J. Biol. Chem.* 276, 9478–9485. doi: 10.1074/jbc.M010501200
- Djordjevic, M., Djordjevic, M., and Severinov, K. (2012). CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs. *Biol. Direct* 7:24. doi: 10.1186/1745-6150-7-24
- Hill, T. L. (2013). *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*. New York, NY: Springer Science & Business Media.
- Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Proc. Natl. Acad. Sci. U.S.A.* 371:20150496. doi: 10.1098/rstb.2015.0496
- Kruger, D. H., and Schroeder, C. (1981). Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. *Microbiol. Rev.* 45, 9–51.
- Kunne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I., Miellet, W. R., Klein, M., et al. (2016). Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell* 63, 852–864. doi: 10.1016/j.molcel.2016.07.011
- Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* 1:7. doi: 10.1186/1745-6150-1-7
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Medina-Aparicio, L., Rebollar-Flores, J., Gallego-Hernandez, A., Vazquez, A., Olvera, L., Gutierrez-Rios, R., et al. (2011). The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi. *J. Bacteriol.* 193, 2396–2407. doi: 10.1128/JB.01480-10
- Mojica, F. J., and Diez-Villasenor, C. (2010). The on-off switch of CRISPR immunity against phages in *Escherichia coli*. *Mol. Microbiol.* 77, 1341–1345. doi: 10.1111/j.1365-2958.2010.07326.x
- Morozova, N., Sabantsev, A., Bogdanova, E., Fedorova, Y., Maikova, A., Vedyaykin, A., et al. (2015). Temporal dynamics of methyltransferase and restriction endonuclease accumulation in individual cells after introducing a restriction-modification system. *Nucleic Acids Res.* 44, 790–800. doi: 10.1093/nar/gkv1490
- Mruk, I., and Blumenthal, R. M. (2008). Real-time kinetics of restriction-modification gene expression after entry into a new host cell. *Nucleic Acids Res.* 36, 2581–2593. doi: 10.1093/nar/gkn097
- Mruk, I., and Kobayashi, I. (2013). To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.* 42, 70–86. doi: 10.1093/nar/gkt711
- Musharova, O., Klimuk, E., Datsenko, K. A., Metlitskaya, A., Logacheva, M., Semenova, E., et al. (2017). Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* 45, 3297–3307. doi: 10.1093/nar/gkx097
- Patterson, A. G., Yevstigneyeva, M. S., and Fineran, P. C. (2017). Regulation of CRISPR-Cas adaptive immune systems. *Curr. Opin. Microbiol.* 37, 1–7. doi: 10.1016/j.mib.2017.02.004
- Pingoud, A., Wilson, G. G., and Wende, W. (2014). Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* 42, 7489–7527. doi: 10.1093/nar/gku447
- Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L., et al. (2010). Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* 77, 1367–1379. doi: 10.1111/j.1365-2958.2010.07265.x
- Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* 75, 1495–1512. doi: 10.1111/j.1365-2958.2010.07073.x
- Ratner, H. K., Sampson, T. R., and Weiss, D. S. (2015). I can see CRISPR now, even when phage are gone: a view on alternative CRISPR-Cas functions from the prokaryotic envelope. *Curr. Opin. Infect. Dis.* 28:267. doi: 10.1097/QCO.000000000000154
- Richter, C., Chang, J. T., and Fineran, P. C. (2012). Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses* 4, 2291–2311. doi: 10.3390/v4102291
- Rodic, A., Blagojevic, B., Zdobnov, E., Djordjevic, M., and Djordjevic, M. (2017). Understanding key features of bacterial restriction-modification systems through quantitative modeling. *BMC Syst. Biol.* 11, 1–15. doi: 10.1186/s12918-016-0377-x
- Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K. A., et al. (2016). Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type IIE CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7626–7631. doi: 10.1073/pnas.1602639113
- Severinov, K., Ispolatov, I., and Semenova, E. (2016). The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Proc. Natl. Acad. Sci. U.S.A.* 3:45. doi: 10.3389/fmolb.2016.00045
- Sternberg, S. H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. *Mol. Cell* 61, 797–808. doi: 10.1016/j.molcel.2016.01.030
- Stratmann, T., Pul, U., Wurm, R., Wagner, R., and Schnetz, K. (2012). RcsB-BglJ activates the *Escherichia coli* leuO gene, encoding an H-NS antagonist and pleiotropic regulator of virulence determinants. *Mol. Microbiol.* 83, 1109–1123. doi: 10.1111/j.1365-2958.2012.07993.x
- Tao, T., Bourne, J. C., and Blumenthal, R. M. (1991). A family of regulatory genes associated with type II restriction-modification systems. *J. Bacteriol.* 173, 1367–1375. doi: 10.1128/jb.173.4.1367-1375.1991
- Touchon, M., Charpentier, S., Clermont, O., Rocha, E. P., Denamur, E., and Branger, C. (2011). CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J. Bacteriol.* 193, 2460–2467. doi: 10.1128/JB.01307-10
- Vijesurier, R. M., Carlock, L., Blumenthal, R. M., and Dunbar, J. C. (2000). Role and mechanism of action of C · PvuII, a regulatory protein

- conserved among restriction-modification systems. *J. Bacteriol.* 182, 477–487. doi: 10.1128/JB.182.2.477-487.2000
- Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* 77, 1380–1393. doi: 10.1111/j.1365-2958.2010.07315.x
- Yang, C. D., Chen, Y. H., Huang, H. Y., Huang, H. D., and Tseng, C. P. (2014). CRP represses the CRISPR/Cas system in *Escherichia coli*: evidence that endogenous CRISPR spacers impede phage P1 replication. *Mol. Microbiol.* 92, 1072–1091. doi: 10.1111/mmi.12614

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Rodic, Blagojevic, Djordjevic, Severinov and Djordjevic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Ribonucleotide Reductases from Bifidobacteria Contain Multiple Conserved Indels Distinguishing Them from All Other Organisms: *In Silico* Analysis of the Possible Role of a 43 aa Bifidobacteria-Specific Insert in the Class III RNR Homolog

Seema Alnajar, Bijendra Khadka and Radhey S. Gupta\*

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Paloma López,  
Centro de Investigaciones Biológicas  
(CSIC), Spain  
Paul Meyers,  
University of Cape Town, South Africa

### \*Correspondence:

Radhey S. Gupta  
gupta@mcmaster.ca

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 17 April 2017

**Accepted:** 11 July 2017

**Published:** 31 July 2017

### Citation:

Alnajar S, Khadka B and Gupta RS  
(2017) Ribonucleotide Reductases  
from Bifidobacteria Contain Multiple  
Conserved Indels Distinguishing  
Them from All Other Organisms: *In*  
*Silico* Analysis of the Possible Role  
of a 43 aa Bifidobacteria-Specific  
Insert in the Class III RNR Homolog.  
*Front. Microbiol.* 8:1409.  
doi: 10.3389/fmicb.2017.01409

Bifidobacteria comprises an important group/order of bacteria whose members have widespread usage in the food and health industry due to their health-promoting activity in the human gastrointestinal tract. However, little is known about the underlying molecular properties that are responsible for the probiotic effects of these bacteria. The enzyme ribonucleotide reductase (RNR) plays a key role in all organisms by reducing nucleoside di- or tri- phosphates into corresponding deoxyribose derivatives required for DNA synthesis, and RNR homologs belonging to classes I and III are present in either most or all *Bifidobacteriales*. Comparative analyses of these RNR homologs have identified several novel sequence features in the forms of conserved signature indels (CSIs) that are exclusively found in bifidobacterial RNRs. Specifically, in the large subunit of the aerobic class Ib RNR, three CSIs have been identified that are uniquely found in the *Bifidobacteriales* homologs. Similarly, the large subunit of the anaerobic class III RNR contains five CSIs that are also distinctive characteristics of bifidobacteria. Phylogenetic analyses indicate that these CSIs were introduced in a common ancestor of the *Bifidobacteriales* and retained by all descendants, likely due to their conferring advantageous functional roles. The identified CSIs in the bifidobacterial RNR homologs provide useful tools for further exploration of the novel functional aspects of these important enzymes that are exclusive to these bacteria. We also report here the results of homology modeling studies, which indicate that most of the bifidobacteria-specific CSIs are located within the surface loops of the RNRs, and of these, a large 43 amino acid insert in the class III RNR homolog forms an extension of the allosteric regulatory site known to be essential for protein function. Preliminary docking studies suggest that this large CSI may be playing a role in enhancing the stability of the RNR dimer complex. The possible significances of the identified CSIs, as well as the distribution of RNR homologs in the *Bifidobacteriales*, are discussed.

**Keywords:** novel features of ribonucleotide reductases, probiotic bacteria, *Bifidobacteriales*, conserved signature inserts and deletions, homology modeling and protein docking studies, extended allosteric site, phylogenetic analysis

## INTRODUCTION

The *Bifidobacteriales* constitute an important order of bacteria within the phylum Actinobacteria (Ventura et al., 2007; Zhi et al., 2009; Gao and Gupta, 2012). While some species belonging to this order are pathogenic (Smith et al., 1992; Bradshaw et al., 2006; Alves et al., 2014; Kenyon and Osbak, 2014) many *Bifidobacteriales* species belonging to the genus *Bifidobacterium* are known for their beneficial health-promoting effects in humans and other mammals (Gibson et al., 1995; Leahy et al., 2005; Masco et al., 2005; Ventura et al., 2009; Cronin et al., 2011). These probiotic bifidobacteria form a significant constituent in the microbiota of the human colon, and exert their effects as commensal microorganisms (Biavati et al., 2000; Turrioni et al., 2008, 2009; Mills et al., 2011; Milani et al., 2014; Ventura et al., 2014). As a result, these bacteria are frequently exploited by the food industry to create consumable products that increase their relative proportion in the gut (Gibson et al., 1995; Sanders, 1998; Masco et al., 2005; Oberg et al., 2011; Ventura et al., 2014). Bifidobacteria are Gram-positive, anaerobic, saccharolytic organisms with a unique metabolic pathway known as the “bifid shunt” (Palframan et al., 2003; Biavati and Mattarelli, 2006; Milani et al., 2015). While many characteristics are known about this important group of bacteria, the biochemical and molecular properties contributing toward their probiotic effects, and adaptability in their respective environments, remain elusive (Ventura et al., 2009; Turrioni et al., 2014).

The present study focuses on the enzyme ribonucleotide reductase (RNR), the sole enzyme capable of reducing nucleoside di- or tri- phosphates (NDPs or NTPs) into deoxyribonucleotides (dNDPs or dNTPs) (Eklund et al., 2001; Nordlund and Reichard, 2006; Torrents, 2014). There are currently three recognized classes of RNRs, named classes I, II, and III, sharing no more than 10% sequence identity across their lengths, which are distributed in different organisms (Logan et al., 1999; Sintchak et al., 2002; Torrents et al., 2002). Class I RNR is further divided into three subclasses viz. Ia, Ib, Ic (Jordan et al., 1996; Jiang et al., 2007; Bollinger et al., 2008). The distributions of these different classes of RNRs within the bacterial domain does not follow any specific pattern that can be correlated with the phylogenies of the bacterial phyla (Torrents et al., 2002; Lundin et al., 2009). However, since the different classes of RNR employ different mechanisms of action and require differing environmental prerequisites to function, we explore their distribution in bifidobacteria in an attempt to identify any unique characteristics that may distinguish them.

Each RNR is capable of reducing all four ribonucleotides into their corresponding deoxyribonucleotides by exhibiting a tightly regulated allosteric substrate specificity site, and employing a convoluted mechanism involving radical chemistry that ultimately results in the removal of a hydrogen from the 3' carbon of the substrate (Brown and Reichard, 1969; Reichard, 1993, 2010; Eriksson et al., 1997; Eliasson et al., 1999). Some RNRs have an additional overall activity site, made possible by the existence of an ATP cone domain at the N-terminus (Thelander and Reichard, 1979). Class I RNRs use NDPs as their substrate, and are aerobic tetramers consisting of one large (R1) and one

small (R2) homodimers (Nordlund and Reichard, 2006). The R2 dimer harbors a dinuclear metallocofactor where the radical is formed and subsequently transferred to the active site located at the R1 subunit. Classes Ia, Ib, and Ic differ in the type of metallocofactor in R2 (manganese and/or iron), as well as the different cofactors required for enzymatic function (Petersson et al., 1980; Nordlund and Reichard, 2006; Jiang et al., 2007; Bollinger et al., 2008; Torrents, 2014). Most bifidobacteria harbor a class Ib RNR (Lundin et al., 2009), whose large and small subunit are encoded by the *nrdE* and *nrdF* genes, respectively, and require NrdH as a reductant and NrdI as a cofactor; in contrast, classes Ia and Ic utilize thioredoxin and/or glutaredoxin as reductants, do not require additional cofactors, and their large and small subunits are encoded by *nrdA* and *nrdB*, respectively (Jordan et al., 1997; Cotruvo and Stubbe, 2008; Roca et al., 2008; Crona et al., 2011). Class II RNRs are not oxygen sensitive, use either NDPs or NTPs as their substrate, and are the only monomeric class of RNR (Tamao and Blakley, 1973; Larsson et al., 2010), however, their structural topology mimics a dimer (Sintchak et al., 2002). No known bifidobacteria harbor a class II homolog, but all bifidobacteria possess a class III RNR. Class III RNRs are encoded by *nrdD* and *nrdG* genes and function under strictly anaerobic conditions, with NTPs as their sole substrates (Garriga et al., 1996; Torrents et al., 2001). They consist of a large R1 subunit (NrdD) that is a dimer in its native state, and works concomitantly with a small activase (NrdG) which generates the radical utilizing a [4Fe-4S] cluster (Eliasson et al., 1992; Sun et al., 1995; Logan et al., 1999). This is unlike class I RNRs, where radical formation by the small subunit is required to induce dimer formation of the large subunit (Ollagnier et al., 1996; Torrents, 2014). Despite the described differences in the properties of the different classes of RNRs, the remarkable structural similarities seen across the three main RNRs strongly suggest a common evolutionary origin of them (Poole et al., 2002; Sintchak et al., 2002; Torrents et al., 2002). In all three RNRs, allosteric regulation involves binding of the dNDP/dNTP products at a 4-helix bundle, involving two helices from each monomeric subunit, at the dimer interface of the enzyme (Uhlen and Eklund, 1994; Larsson et al., 2001). The allosteric regulation causes conformational changes at a highly conserved 10 stranded  $\alpha/\beta$  barrel where the active site “finger loop” structure resides in its center, or is brought to its center upon activation (Aurelius et al., 2015).

Although previous studies have significantly contributed to the current understanding of the structure and function of the different RNRs, in the present work we focus on the specific biochemical/molecular properties of the RNRs from bifidobacteria that may shed light on their unique physiological effects. Our earlier work describes a number of conserved signature indels (CSIs) in the homologs of many important proteins from that are uniquely found in all *Bifidobacteriales* (Zhang et al., 2016). These CSIs represent vertically transferred genetic changes that are indicated to have occurred in a common ancestor of the group in which they are found, thus asserting their value as highly specific molecular markers. In the present work, we have performed similar comparative genomic studies that have led to the identification of several novel CSIs in class Ib



and III RNR homologs that are shared by all genome-sequenced *Bifidobacteriales* species that contain the respective protein homolog(s), but are absent in all other bacteria. We also describe the results of protein modeling which illustrate the structural location of these CSIs, as well as the results of preliminary *in silico* docking studies which suggest that one of the large CSIs [a 43 amino acid (aa) insertion in class III RNR] may be playing a role in NrdD complex stability.

## MATERIALS AND METHODS

### Identification of Conserved Signature Indels

The approach used to identify CSIs in RNR was as described in earlier work (Gupta, 2014; Zhang et al., 2016). Multiple sequence alignments (MSAs) were initially created using the Clustal\_X 2.1 (Larkin et al., 2007; Goujon et al., 2010) program for the protein sequences of NrdE, NrdF, NrdH, NrdD, and NrdG homologs from about 10–15 *Bifidobacteriales* species, as well as 8–10 species from other groups/phyla of bacteria. These sequence alignments were examined for the presence of conserved indels that are limited to the *Bifidobacteriales* homologs and are flanked on both sides by at least five conserved residues in the neighboring 30–40 aa. A detailed Blastp search (Altschul et al., 1997) was then conducted on the sequence region containing the potential conserved indels to investigate the species-specificities of the identified indels. The indels that were not flanked by conserved regions were not further investigated in our work. The signature files shown here were created using SIG\_CREATE and SIG\_STYLE from the GLEANS.net program as described in earlier work (Gupta, 2014; Zhang et al., 2016). Unless otherwise indicated, all of the reported CSIs are specific for the *Bifidobacteriales* homologs and similar CSIs were not observed in homologs from any other bacterial species within the top 500–1000 blast hits examined.

### Phylogenetic Tree Construction

In this study we have constructed three separate phylogenetic trees: (i) based on NrdE (large subunit of class Ib RNR) sequences, (ii) based on NrdD (large subunit of class III RNR) sequences, and (iii) based on the large subunit sequences from class I (NrdA, NrdE), II (NrdJ), and III (NrdD) RNRs. For these studies, NrdE and NrdD homologs from all genome sequenced bifidobacterial species were obtained from the NCBI GenBank sequence database (Benson et al., 2017). The species represented in the tree based on NrdD sequences included 49 of 58 validly published *Bifidobacterium* species, the two known *Scardovia* species, all three *Alloscardovia* species, and the single species known from the *Parascardovia* and *Gardnerella* genera. The tree based on NrdE sequences similarly included the subset of these *Bifidobacteriales* species where the protein homolog was detected. Sequences from members of the *Bifidobacteriales* genera *Aeriscardovia* and *Pseudoscardovia* were not available at the present time and were not included in our study. For each tree, a MSA of RNR homologs was created using the Clustal\_X 2.1 (Larkin et al., 2007; Goujon et al., 2010) program. For each of

these trees, we have additionally included a number of outgroup species (20 species for the NrdD tree, 23 species for the NrdE tree) from other orders in the Actinobacteria phylum, as well as Firmicutes species. For the tree concerning the sequences from large subunits of all RNR classes, we have included <10 NrdE and <10 NrdD sequences from representative *Bifidobacteriales*, in addition to several species across various bacterial phyla in order to depict the evolutionary history of RNR classes. The MEGA 6 program (Tamura et al., 2013) was used to construct a maximum likelihood (ML) tree based on 1000 bootstrap replicates for each alignment employing the Whelan and Goldman model substitution method (Whelan and Goldman, 2001). Gaps and regions with missing data from the sequence alignments were completely removed. In each case, a discrete Gamma distribution was used to model evolutionary rate differences among sites (five categories) and the Jones–Taylor–Thornton substitution method was used to compute the initial trees for the heuristic search using the Neighbor-joining method with a matrix of pairwise distances (Jones et al., 1992).

### Homology Modeling of RNR Homologs and Structural Analysis of CSIs

The approach used to model the CSIs involves homology modeling based on previously crystallized class Ib and III RNR proteins. A Position-Specific-Iterated Blastp search (Altschul et al., 1997) was performed on *Bifidobacterium longum* NrdE (Accession no. EPE39971) and NrdD (Accession no. KXS29127) sequences against the PDB database which revealed that the class Ib RNR from *Salmonella typhimurium* (PDB ID: 1PEQ) (Uppsten et al., 2003) and class III RNR from Enterobacteria phage T4 (PDB ID: 1H7B) (Larsson et al., 2001) exhibited the highest sequence similarity to the *Bifidobacteriales* homologs and provided suitable templates for homology modeling of the RNR isoforms of *B. longum*. A conserved domain search (CD-Search) (Marchler-Bauer and Bryant, 2004) was conducted on the *B. longum* sequences. Homology modeling was performed using MODELLER v9.11 (Eswar et al., 2007) and the top 500 models were initially created and ranked on the basis of their Discrete Optimized Protein Energy (DOPE) scores (Shen and Salí, 2006). The selected models of RNR homologs with the highest DOPE score were then submitted to the GalaxyRefine server (Heo et al., 2013; Lee et al., 2016) to obtain atomic-level energy minimization and to improve the stereochemical quality of the model. The secondary structure elements in the regions containing CSIs were examined and compared with results of the PSIPRED and CONCORD analyses to ensure their reliability (Jones, 1999; Wei et al., 2011; Buchan et al., 2013). The stereochemical properties of the final models were assessed using four independent servers: RAMPAGE, ERRAT, PROSA and Verify3D (Bowie et al., 1991; Luthy et al., 1992; Colovos and Yeates, 1993; Sippl et al., 1999; Lovell et al., 2003; Wiederstein and Sippl, 2007). These validation tools utilize a dataset of highly refined solved structures to evaluate the statistical significance of models based on the conformation, location, and the environment of individual amino acids in the protein sequence, as well as the model's overall structural stability. The structural alignments of the models with the respective templates were carried out using PyMOL

**TABLE 1** | Results of protein–protein docking studies for the *Bifidobacterium longum* RNR (NrdD subunit) either containing or lacking the large 43 aa CSI.

Docking server	Protein models of the NrdD		
	CSI-containing (Extended helix)	CSI-containing (based on PSIPRED/CONCORD)	CSI lacking
ClusPro	–1297.200	–1190.900	–953.000
ClusPro + ROSIE	–1092.851	–1110.039	–1042.091
PatchDock	*	16192.000	11234.000
PatchDock + ROSIE	*	–1109.473	–1033.714
“ROSIE”	–1105.550	–1114.689	–1043.235

The results of docking studies are shown for two different models of the CSI-containing protein (Extended helix and based on PSIPRED/CONCORD) and compared with those for the model for CSI-lacking protein. The monomers based on the indicated models were submitted to ClusPro and PatchDock servers. Results from the ClusPro and PatchDock servers were also submitted to the ROSIE (RosettaDock) server for local refinement. The results shown under the heading “ROSIE” were obtained by submitting model monomers that were superimposed onto the crystallized 1H7B (Larsson et al., 2001) template dimer. For the ClusPro and ROSIE servers, the lower binding energy (more negative values) is indicative of the increase in stability of the docking complexes, while for the PatchDock server higher (positive) scores indicate stronger binding affinity (see Materials and Methods section for further details). Asterisks (\*) indicate that no biologically relevant complexes were obtained from the indicated docking servers.

Version 1.8 (Schrödinger, 2016) in order to analyze the location and the structural features of the CSIs in the protein structure. This procedure was followed to create the homology models of both of the RNR homologs found in bifidobacteria. In addition, a structural model of class III RNR was also generated using I-TASSER, an online server that uses threading to predict three dimensional protein structure (Zhang, 2008; Roy et al., 2010; Yang et al., 2015).

## Protein–Protein Docking Analysis of the Class III RNR Homologs

Protein–protein docking studies were performed in order to assess the possible role of a large CSI in the formation or stabilization of the dimeric structure of class III RNR in bifidobacteria. A structural model of the class III RNR from *B. longum* was created by removing the CSI residues from its primary sequence, using the methods described above for other RNR homologs. An additional structural model of RNR was generated with the CSI region constructed as an extended helix. In this structural model, the CSI has a slightly different secondary structure than those of the models that followed PSIPRED/CONCORD, or I-TASSER predictions. This was done in an attempt to be inclusive of multiple possible structural conformations of the CSI. Four structures of the anaerobic RNR monomer viz. PSIPRED/CONCORD based model, I-TASSER generated model, model with extended helix, and the CSI-lacking model, were submitted to two independent web-based protein–protein docking programs using default parameters, viz. PatchDock Version B 1.3 (Schneidman-Duhovny et al., 2005) and ClusPro Version 2.0 (Comeau et al., 2004). PatchDock is an efficient molecular docking algorithm that employs a geometry-based shape complementarity approach which aims

to yield refined atomic contacts of protein–protein complexes. Its scoring function takes into consideration both geometric fit and atomic desolvation energy (Duhovny et al., 2002). On the other hand, ClusPro utilizes PIPER, a rigid body docking program, which is based on a novel Fast-Fourier Transform (FFT) docking approach with pairwise potential. Its scoring function is thus based on pairwise interaction potentials (Comeau et al., 2004; Kozakov et al., 2006). The resulting top scoring dimer complex models of RNR from each server (if any) were then refined using the RosettaDock (ROSIE) server (Lyskov and Gray, 2008; Chaudhury et al., 2011; Lyskov et al., 2013). For the docking scores of ClusPro and RosettaDock, the lower (negative) binding energy value indicates improved stability of the docking complexes. In the case of PatchDock, the geometry shape complementarity score was utilized to determine rank, and higher (positive) scores indicate stronger binding affinity (see also notes in Table 1). In addition, the monomeric forms of each of the four models were structurally aligned with the established biological assembly of the RNR dimer, and the resulting dimer orientations were utilized as additional inputs for submission to the ROSIE server. The resulting refined structure from ROSIE with the lowest total score, maximum cluster size and the smallest RMSD with respect to the solved structure of RNR complex, was chosen as a representative structure for detail interface interaction analysis. To analyze the dimer interface, this class III RNR dimeric output structure was submitted to the PDBePISA Version 1.48 server, using default parameters (Krissinel and Henrick, 2007).

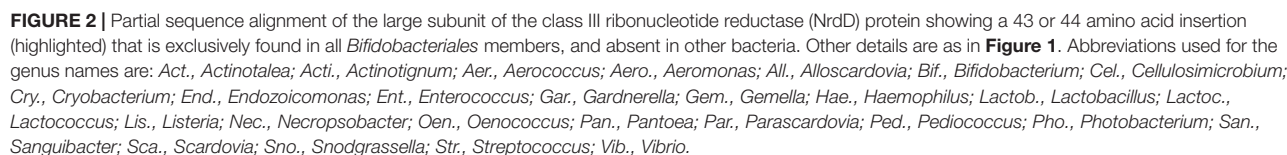
## RESULTS

### Identification of Conserved Signature Indels in Class I and Class III RNR Homologs and Their Phylogenetic Implications

Comparative analysis of the *Bifidobacteriales* genomes indicated that all of the sequenced species from this order contain an anaerobic class III RNR homolog. In addition, an aerobic class I RNR belonging to the class Ib group was also found in most species from this order except *B. adolescentis*, *B. angulatum*, *B. dentium*, *B. gallicum*, *B. cuniculi*, *B. lemurum*, *B. merycicum*, *B. moukalabense*, *B. ruminantium*, and members of the genera *Parascardovia* and *Scardovia*. The sequences of class Ib and III RNRs were examined for the presence of any CSIs that are specific for bifidobacteria. The results of these studies have identified three CSIs in the large subunit of the class Ib RNR (NrdE) homologs, which are specifically found in the bifidobacterial enzyme. Sequence information for two of these CSIs, which are comprised of 4 and 2 aa inserts in a conserved region of the NrdE protein, is shown in Figure 1. As seen in the figure, both of these CSIs are flanked on either side by conserved regions and while they are commonly shared by all of the bifidobacteria harboring the NrdE homolog, they are not present in any other bacterial species in the top 500 blast hits. Sequence information for one additional CSI in the NrdE protein, consisting of a 1 aa



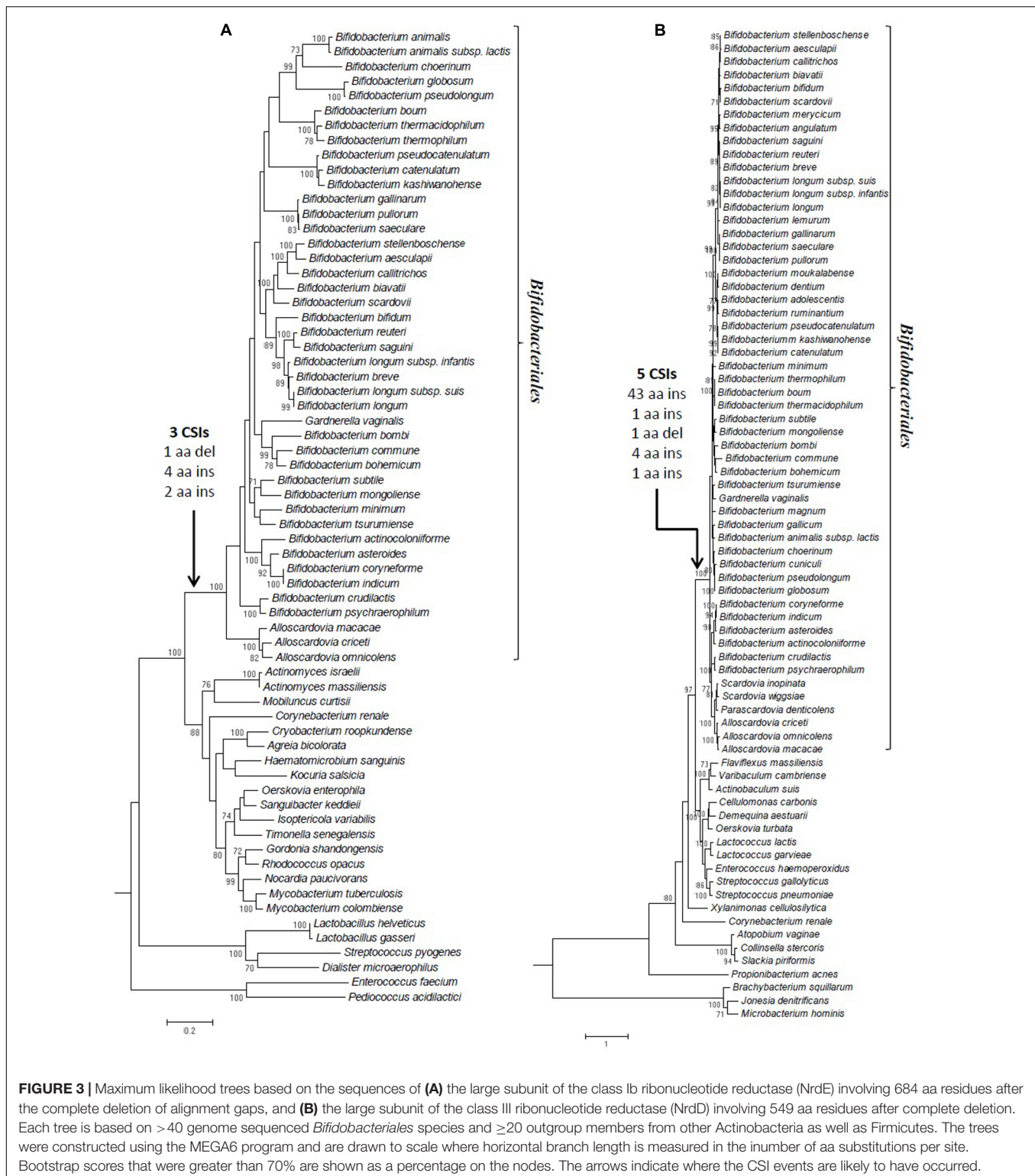
**Figure 1** | Partial sequence alignment of the large subunit of the class Ib ribonucleotide reductase (NrdE) protein showing two conserved inserts (highlighted) that are exclusively found in all *Bifidobacteriales* members that carry the homolog, but absent in other bacteria. The dashes in the sequence alignment denote identity with the amino acid found on the top line. The Genbank accession numbers of the sequences are shown in the second column. The results are shown for only a limited number of species, however, other species not shown showed similar pattern as described here. Abbreviations used for the genus names are: *All.*, *Alloscardovia*; *Art.*, *Arthrobacter*; *Bif.*, *Bifidobacterium*; *Bren.*, *Brenneria*; *Brev.*, *Brevibacterium*; *Cell.*, *Cellulomonas*; *Cry.*, *Cryobacterium*; *Gar.*, *Gardnerella*; *Gor.*, *Gordonia*; *Jon.*, *Jonesia*; *Mob.*, *Mobiluncus*; *Myc.*, *Mycobacterium*; *Noc.*, *Nocardia*; *Rou.*, *Rouxiella*; *Seg.*, *Segniliparus*; *Ser.*, *Serratia*; *Vib.*, *Vibrio*; *Wil.*, *Williamsia*; *Yer.*, *Yersinia*.



of them also play important role in conjunction with the CSIs in the novel functional aspect(s) of the RNRs from bifidobacteria.

Maximum-likelihood phylogenetic trees were constructed for the class Ib and class III RNRs protein sequences based on the NrdE and NrdD proteins, and these trees are shown in **Figures 3A,B**, respectively. In addition to the sequences from a large number of *Bifidobacteriales* species covering the order, the tree also contains information for several other Actinobacteria as well as a limited number of Firmicutes species; the sequences





from the Firmicutes species were used to root the trees. The sequences from *Bifidobacteriales* species formed strongly supported monophyletic clades in both trees. Because the sequence alignments used for construction of these phylogenetic

trees did not contain any sequence gaps, the observed branching pattern was not influenced by the presence of the identified CSIs. Therefore, the distinct branching of bifidobacteria observed in both trees supports the notion that the reported CSIs in the

NrdD and NrdE proteins most likely first occurred in a common ancestor of the order *Bifidobacteriales*, and were inherited by descendants due to incurring an evolutionary advantage. In addition to these trees, we have also constructed a tree based on the sequences of the large subunit from the three main RNR classes (Supplementary Figure 8). The three classes of RNR formed distinct clades in the tree, which were separated from each other by long branches. Based on the midpoint rooting of the tree, the sequences from the class III RNR, which function under strictly anaerobic conditions, were found to form a sister clade to sequences from the classes I and II. The observed branching of the class III RNR in the tree is in agreement with earlier work (Reichard, 1993; Logan et al., 1999; Larsson et al., 2001; Poole et al., 2002; Sintchak et al., 2002; Torrents et al., 2002) suggesting that this class of RNR represents the ancestral form of the reductase. Although phylogenetic analysis can shed light on the evolutionary history of the three types of RNR, it does not explain the variable distribution of these classes in different organisms. As an important protein, there is at least one type of RNR in every organism. However, the combination of different RNRs which are found in various organisms is unpredictable and it does not show any correlation with the evolutionary histories of the organisms (Torrents et al., 2000; Lundin et al., 2009, 2010; Torrents, 2014).

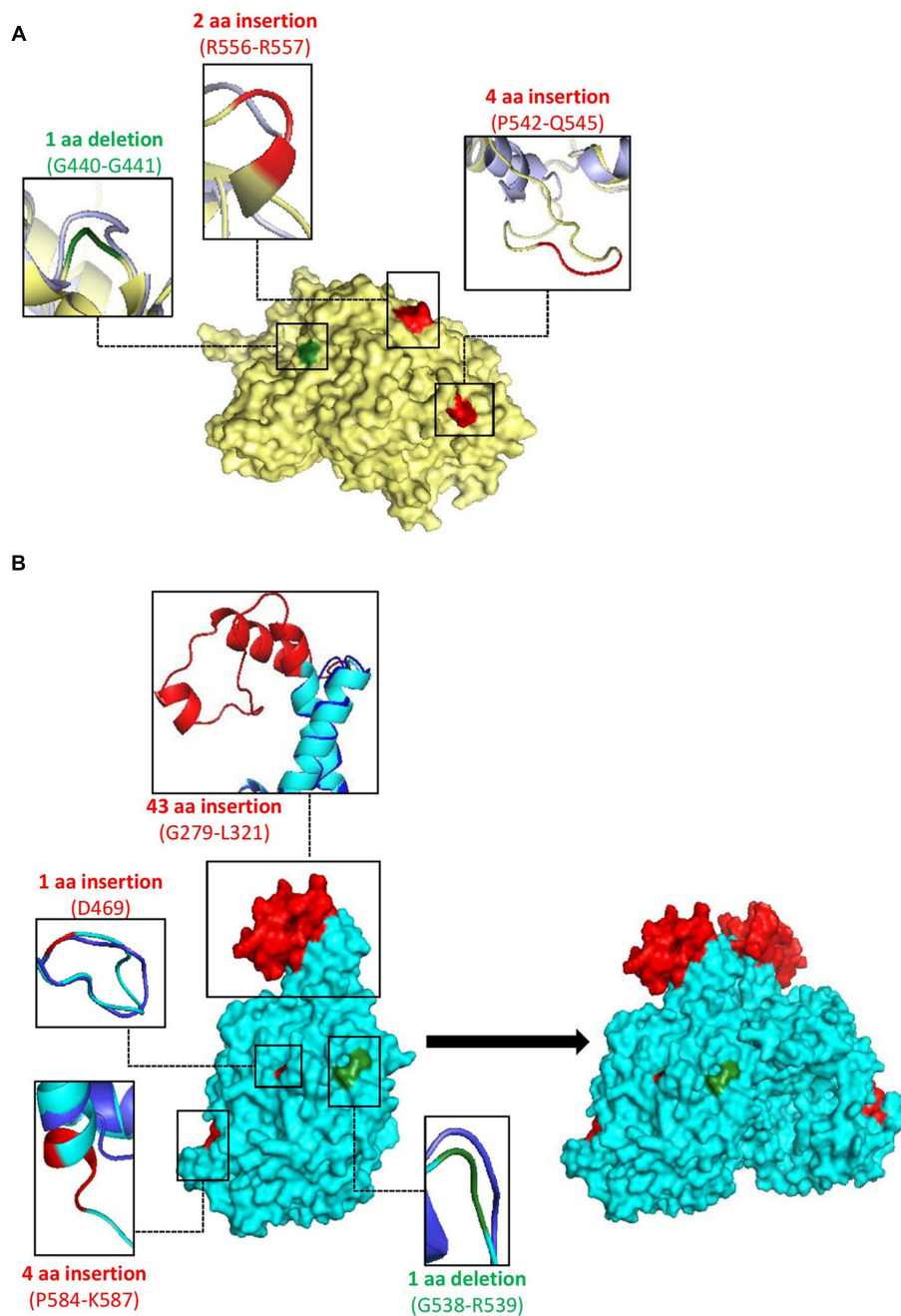
## Locations of the CSIs in the Ribonucleotide Reductase Homologs Structures

To gain insights into the possible significance of the identified CSIs, homology models for the class Ib and III RNRs from *B. longum* were constructed (see Materials and Methods section) based on previously crystallized template structures of class Ib and III RNR proteins (Larsson et al., 2001; Uppsten et al., 2003). After the validation of the homology models using a variety of tools described under section “Materials and Methods,” a superimposition of the final selected models with the template structures was carried out using PyMOL to determine the locations of the identified CSIs in the structures of the class Ib and III proteins. The locations of the three CSIs identified in class Ib RNR in the modeled structure of the NrdE protein is shown in **Figure 4A**. As seen, all three CSIs in the NrdE homolog were located within the surface loops of the protein (**Figure 4A**). However, of these CSIs, the 2 aa insert also appears to extend a helix. The current model is in agreement with secondary structure analyses (PSIPRED/CONCORD), and yielded reassuring measurements by ERRAT, Verify3D, RAMPAGE and PROSA. The locations of these CSIs in the structure of the *B. longum* NrdE subunit indicate that they are topologically distant from the dimer interface/allosteric regulatory site, as well as the active site as seen in Supplementary Figure 9A (Uppsten et al., 2003, 2006). However, the locations of these CSIs within surface loops in the NrdE structure indicate that they could be involved in mediating novel protein–protein interactions (Cherkasov et al., 2006; Singh and Gupta, 2009; Gupta, 2016; Zhang et al., 2016; Khadka and Gupta, 2017).

Of the five CSIs found in the NrdD homologs (Class III) of bifidobacteria, the structural locations of four indels could be determined and are illustrated on the modeled structure (**Figure 4B**). Structural location of one of the CSIs present near the C-terminal end could not be determined as the structural information for the corresponding region was absent from the template structure (PDB ID: 1H7B) used for homology modeling (Larsson et al., 2001). Similar to the indels found in the NrdE structure, most of the CSIs in the NrdD structure are also found on surface loops that are structurally distant from the active site (**Figure 4B** and Supplementary Figure 9B). The 4 aa insert, also located on the surface exposed region, appears to form a loop and elongate a helix. The large 43 aa insert in NrdD exists as an elongation of the allosteric regulatory region, in between two helices that form the 4-helix bundle in the NrdD dimer (Uhlen and Eklund, 1994; Logan et al., 1999; Larsson et al., 2001). Since the 43 aa insert did not correspond to a characterized domain or motif, elucidating its structure was a challenging task. We present a model that is in agreement with secondary structure predictions from the *B. longum* primary sequence (PSIPRED/CONCORD) (**Figure 4B**), and an additional model according to the prediction made *in silico* by the I-TASSER server (Supplementary Figure 10B). As seen in Supplementary Figure 10B, in the I-TASSER model, the 43 aa insert appears to form two helices that are connected to one another by a loop, and each are also connected to the two existing helices by loops. The orientation of the insert is such that it folds back toward the bulk of the protein, and the dimer interface is relatively uninfluenced. In the case of the model generated based on the secondary structure predictions, the CSI forms an extension of the two helices, along with two small helices connected by loops in between (**Figure 4B**). In order to be inclusive of all reasonable possibilities, we also modeled the NrdD homolog based on the hypothesis that, instead of small helices with breaks induced by loops according to the PSIPRED/CONCORD model, perhaps the loops connecting the main helices to the small helices may be extended helices without loop-induced breaks (Supplementary Figure 10A). This extended helix hypothesis is a corollary to the observation that the elongated helix in class III compared to class I has important functional significance regarding allosteric binding, and also influences dimer packing (Larsson et al., 2001). All three models were refined and validation scores were maximized in these CSI-containing regions.

## Analyzing the Possible Functional Significance of the Large Conserved Insert

To determine the possible role of the large 43 aa insert in dimer formation or complex stability, we have performed a series of docking studies to reveal dimerization potentials of the models compared to an additional model of bifidobacterial NrdD that lacks the CSI. These four models (I-TASSER, PSIPRED/CONCORD, extended helix hypothesis and CSI-lacking) were submitted to two online servers viz. ClusPro and PatchDock. The complexes obtained from ClusPro and PatchDock were refined and scored using ROSIE, and an



**FIGURE 4 |** Surface representations of homology models of **(A)** the class Ib ribonucleotide reductase monomer from *Bifidobacterium longum*, modeled from the 1PEQ template (Uppsten et al., 2003); the cartoon representations are of the model (yellow) superimposed on the template (purple) and **(B)** the class III ribonucleotide reductase monomer (NrDd) from *B. longum*, modeled from the 1H7B template (Larsson et al., 2001), with secondary structure in agreement with PSIPRED and CONCORD analyses; the cartoon representations are of the model (cyan) superimposed on the template (dark blue). The observed dimer on the right of **(B)** is a result of docking analysis from ClusPro, refined by ROSIE. The conserved signature insertions identified in these proteins are highlighted in red, and the positions of the deletions are shown in green.

additional dimerization measurement was performed by superimposing the models with the available experimentally solved structure of the NrDd dimer (Larsson et al., 2001). The docking scores of the dimer complexes from the protein–protein docking studies are summarized in **Table 1**. The PatchDock

server did not yield any dimer complex for the modeled proteins containing extended helix that was plausible with the known biological assembly. However, the docking scores for the CSI-containing protein model based on PSIPRED/CONCORD were consistently improved (or superior) for all servers in



comparison to the protein model lacking the CSI. Thus, the model generated according to the PSIPRED/CONCORD prediction is more likely to approximate the true structure of the class III RNR in *B. longum*. For all docking servers, the I-TASSER generated protein model did not form a plausible dimer complex and hence its results are not shown. This may be due to the fact that in the I-TASSER model of NrdD, the CSI was found to protrude away from the dimeric interface (Supplementary Figure 10B).

To examine if any of the residues from the CSI are involved in dimer interaction, the structural coordinate file for the dimeric form of the PSIPRED/CONCORD model was submitted to the PDBePISA server (Krissinel and Henrick, 2007). Analysis of the results obtained suggests that two residues from the large CSI (viz. Gly279, Met280 for *B. longum* NrdD) are present at the protein dimer interface. However, both of these residues, which are present at the N-terminal end of this large CSI, are not conserved in other *Bifidobacteriales* homologs. Thus, it is difficult to infer with any degree of confidence the possible role of these two residues in protein dimerization. Besides from these two residues, the remaining CSI residues are partly or fully solvent accessible near the interface.

## DISCUSSION

The *Bifidobacteriales* are an important group of bacteria that consist of both pathogenic species and health-promoting commensal microorganisms that are frequently exploited in the food industry as probiotics (Gibson et al., 1995; Sanders, 1998; Leahy et al., 2005; Masco et al., 2005; Ventura et al., 2009; Cronin et al., 2011). However, the use of these bacteria as probiotics faces several challenges as very little is understood about the mechanism responsible for the beneficial effects exerted by bifidobacteria (Oberge et al., 2011). In the present work, we have identified several novel signatures in the form of CSIs in the sequences of both classes I and III RNR homologs that differentiate the RNR homologs from *Bifidobacteriales* from all those found in all other organisms. Earlier work on CSIs (including 1–2 amino acid indels) in several important proteins (e.g., GroEL, DnaK, GyrB, PIP5K, etc.) provides evidence that the CSIs play important functional roles in CSI-containing organisms, and deletion or other changes in the CSIs adversely impact cell growth or other critical functions (Chatterji et al., 2000; Singh and Gupta, 2009; Schoeffler et al., 2010; Clarke and Irvine, 2013; Gupta et al., 2017). In this context, the results reported here that both classes I and III RNR homologs harbor multiple CSIs that are uniquely present in all bifidobacteria is of much interest. These CSIs serve to clearly differentiate the bifidobacterial RNR homologs from those found in all other organisms, and they could function in conjunction with each other to impart certain novel functional characteristic(s) that is only shared by the classes I and III RNR homologs from bifidobacteria.

The distribution of RNR homologs in the *Bifidobacteriales* reveals that while all bifidobacteria species contain a class III

anaerobic RNR homolog (NrdDG), the class Ib aerobic RNR homologs (NrdEF) were not found (or detected) in a number of *Bifidobacterium* species (viz. *B. adolescentis*, *B. angulatum*, *B. dentium*, and *B. gallicum*, *B. cuniculi*, *B. lemurum*, *B. merycicum*, *B. moukalabense*, *B. ruminantium*) as well as in members of the genera *Parascardovia* and *Scardovia* (Lundin et al., 2009). The bifidobacteria species lacking the class Ib RNR homologs do not show any specific branching pattern, or belong to any specific clade(s) of bifidobacteria, but they appear to be distributed sporadically within the order *Bifidobacteriales* (Figure 3). Thus, it is likely that the genes for both classes I and III RNR were present in the common ancestor of all *Bifidobacteriales* and subsequently some species have lost the genes for the class I RNR. Since the different identified CSIs in the classes I and III RNR homologs are present in all bifidobacteria, the most likely explanation for this fact is that the genetic changes responsible for the observed CSIs occurred in a common ancestor of the order *Bifidobacteriales*, presumably at the time of divergence of this group of bacteria from other organisms, similar to the CSIs in many other proteins that are uniquely found in the members of this order (Zhang et al., 2016).

The biological significance of the wide-spread presence of an aerobic RNR (class Ib) in bifidobacterial species, which are generally regarded as anaerobic organisms, remains to be understood. It is known that bifidobacterial species exhibit varying levels of aerotolerance which affects their viability outside of their natural habitats (e.g., gastrointestinal tract, mouth and vagina of mammals) upon exposure to an oxidative environment (Biavati and Mattarelli, 2006; Oberge et al., 2011). Under these conditions, RNRs are pivotal in maintaining a pool of dNDPs/dNTPs to compensate for DNA and protein damage by reactive oxygen species (ROS), ultimately alleviating the detrimental effects of superoxide stress. Thus, a reasonable assumption is that the class I aerobic RNR may play an important role in the tolerance of bifidobacteria to oxidative environment. Although how bifidobacteria manage oxidative exposure remains to be understood, in few studies that have explored the gene expression of both NrdEFHI and NrdDG systems in *Bifidobacterium* species, including data available from the Gene Expression Omnibus<sup>1</sup>, reveal that, under oxidative stress, the expression of both gene systems was rapidly induced by the altered environment; first, the class Ib system was upregulated, followed by the class III system and other proteins, including proteolytic enzymes (Larsson et al., 2001; Edgar et al., 2002; Xiao et al., 2011; Zuo and Chen, 2016). However, further work is necessary to understand the biological and physiological roles of the class I RNR in bifidobacteria under normal conditions and during oxidative stress.

Earlier work on CSIs in protein structures provides evidence that most of the studied CSIs are located in the surface loops of proteins (Geszvain et al., 2004; Akiva et al., 2008; Singh and Gupta, 2009; Gupta et al., 2017; Khadka and Gupta, 2017). The results of our modeling of the identified CSIs in the NrdE and NrdD protein structures also show that all of described

<sup>1</sup><http://www.ncbi.nlm.nih.gov/geo/>



CSIs in the classes I and III RNR protein subunits are present on the surface loops of these proteins (**Figure 4**). Based on extensive earlier work, the surface loops in protein structures often serve as platforms for facilitating novel protein–protein or protein–ligand interactions that are specific for the CSI-containing organisms, without affecting the core functions of the target proteins (Geszvain et al., 2004; Akiva et al., 2008; Singh and Gupta, 2009; Schoeffler et al., 2010; Gupta et al., 2017; Khadka and Gupta, 2017). In a number of cases, the surface loops formed by the CSIs have also been shown to play important role in determining the oligomeric state of the proteins (Itzhaki et al., 2006; Akiva et al., 2008; Hashimoto and Panchenko, 2010). Based on these studies, it is expected that the identified CSIs in the RNR homologs of bifidobacteria will also play novel and functionally important roles that are specific for bifidobacteria. There is no information available at present regarding the biochemical properties of the RNR homologs from bifidobacteria or whether they exhibit any novel functional characteristics. However, the identified CSIs provide highly specific tools for the genetic and biochemical exploration of novel functional characteristics of the RNRs from bifidobacteria.

Of the five different CSIs identified within the NrdD homologs, one of these CSIs is a large 43 aa insert, located proximal to the allosteric regulatory site in the protein monomer (**Figure 4B** and Supplementary Figure 9B) (Larsson et al., 2001). This regulatory site consists of a 4-helix bundle where each monomer contributes two helices. These helices are significantly longer in class III RNRs compared to those of class I (~21 aa), resulting in an altered dimer packing, mode of effector binding and consequential conformational changes to the active site (Larsson et al., 2001). Due to the presence of the 43 aa CSI in the class III bifidobacterial RNRs in this region, this helical region is further elongated in bifidobacteria, which is suggestive of a feature uniquely shared by these bacteria. In the absence of any information regarding the structure of the NrdD protein from bifidobacteria, or the conformation of this large CSI in the protein structure, it is difficult to predict the precise role that this CSI may play in the RNR structure and/or function. However, the results of our preliminary *in silico* protein–protein

docking using three separate docking servers suggest that when the CSI is oriented in such a way that it can interact with the other monomer, the stability of the dimer complex is improved in the presence of the 43 aa CSI, in comparison to the similar docking studies carried out with the protein lacking the CSI (**Figure 4B** and **Table 1**). The results from our *in silico* analyses are broadly suggestive of one possible function of this large CSI. However, a clearer understanding of the functional significances of the identified CSIs should emerge from future biochemical and structural studies on classes I and III RNRs. Nevertheless, the results presented here highlight the many unique sequence features of the bifidobacterial RNRs, whose further investigations could provide important insights into novel functional aspects of these enzymes in bifidobacteria.

## AUTHOR CONTRIBUTIONS

SA identified some of the conserved inserts, carried out phylogenetic and structural work and wrote the draft manuscript. BK was involved in the structural analysis of the identified conserved inserts and read and commented on the draft manuscript. RG identified the initial CSIs and conceived and directed the entire project and edited the final version of the manuscript.

## FUNDING

This work was supported by the research grant number 249924 from the Natural Sciences and Engineering Research Council of Canada.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01409/full#supplementary-material>

## REFERENCES

- Akiva, E., Itzhaki, Z., and Margalit, H. (2008). Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13292–13297. doi: 10.1073/pnas.0801207105
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Alves, P., Castro, J., Sousa, C., Cereija, T. B., and Cerca, N. (2014). *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an *in vitro* biofilm formation model. *J. Infect. Dis.* 210, 593–596. doi: 10.1093/infdis/jiu131
- Aurelius, O., Johansson, R., Bagenholm, V., Lundin, D., Tholander, F., Balhuizen, A., et al. (2015). The crystal structure of *Thermotoga maritima* class III ribonucleotide reductase lacks a radical cysteine pre-positioned in the active site. *PLoS ONE* 10:e0128199. doi: 10.1371/journal.pone.0128199
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2017). GenBank. *Nucleic Acids Res.* 45, D37–D42. doi: 10.1093/nar/gkw1070
- Biavati, B., and Mattarelli, P. (2006). “The family *Bifidobacteriaceae*,” in *The Prokaryotes*, eds M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (New York, NY: Springer), 322–382.
- Biavati, B., Vescovo, M., Torriani, S., and Bottazzi, V. (2000). Bifidobacteria: history, ecology, physiology and applications. *Ann. Microbiol.* 50, 117–132.
- Bollinger, J. M. Jr., Jiang, W., Green, M. T., and Krebs, C. (2008). The manganese(IV)/iron(III) cofactor of *Chlamydia trachomatis* ribonucleotide reductase: structure, assembly, radical initiation, and evolution. *Curr. Opin. Struct. Biol.* 18, 650–657. doi: 10.1016/j.sbi.2008.11.007
- Bowie, J. U., Luthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164–170. doi: 10.1126/science.1853201
- Bradshaw, C. S., Tabrizi, S. N., Fairley, C. K., Morton, A. N., Rudland, E., and Garland, S. M. (2006). The association of *Atopobium vaginiae* and *Gardnerella vaginalis* with bacterial vaginosis and recurrence after oral metronidazole therapy. *J. Infect. Dis.* 194, 828–836. doi: 10.1086/506621
- Brown, N. C., and Reichard, P. (1969). Role of effector binding in allosteric control of ribonucleoside diphosphate reductase. *J. Mol. Biol.* 46, 39–55. doi: 10.1016/0022-2836(69)90056-4

- Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013). Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* 41, W349–W357. doi: 10.1093/nar/gkt381
- Chatterji, M., Unniraman, S., Maxwell, A., and Nagaraja, V. (2000). The additional 165 amino acids in the B protein of *Escherichia coli* DNA gyrase have an important role in DNA binding. *J. Biol. Chem.* 275, 22888–22894. doi: 10.1074/jbc.M001047200
- Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., and Gray, J. J. (2011). Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS ONE* 6:e22477. doi: 10.1371/journal.pone.0022477
- Cherkasov, A., Lee, S. J., Nandan, D., and Reiner, N. E. (2006). Large-scale survey for potentially targetable indels in bacterial and protozoan proteins. *Proteins* 62, 371–380. doi: 10.1002/prot.20631
- Clarke, J. H., and Irvine, R. F. (2013). Evolutionarily conserved structural changes in phosphatidylinositol 5-phosphate 4-kinase (PI5P4K) isoforms are responsible for differences in enzyme activity and localization. *Biochem. J.* 454, 49–57. doi: 10.1042/BJ20130488
- Colovos, C., and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2, 1511–1519. doi: 10.1002/pro.5560020916
- Comeau, S. R., Gatchell, D. W., Vajda, S., and Camacho, C. J. (2004). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 20, 45–50. doi: 10.1093/bioinformatics/btg371
- Cotruvo, J. A. Jr., and Stubbe, J. (2008). NrdI, a flavodoxin involved in maintenance of the diferric-tyrosyl radical cofactor in *Escherichia coli* class Ib ribonucleotide reductase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14383–14388. doi: 10.1073/pnas.0807348105
- Crona, M., Torrents, E., Rohr, A. K., Hofer, A., Furrer, E., Tomterm, A. B., et al. (2011). NrdH-redoxin protein mediates high enzyme activity in manganese-reconstituted ribonucleotide reductase from *Bacillus anthracis*. *J. Biol. Chem.* 286, 33053–33060. doi: 10.1074/jbc.M111.278119
- Cronin, M., Ventura, M., Fitzgerald, G. F., and van Sinderen, D. (2011). Progress in genomics, metabolism and biotechnology of bifidobacteria. *Int. J. Food Microbiol.* 149, 4–18. doi: 10.1016/j.ijfoodmicro.2011.01.019
- Duhovny, D., Nussinov, R., and Wolfson, H. J. (2002). “Efficient unbound docking of rigid molecules,” in *Proceedings of the Second International Workshop on Algorithms in Bioinformatics (WABI)* 2452 (Berlin: Springer-Verlag), 185–200. doi: 10.1007/3-540-45784-4\_14
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207
- Eklund, H., Uhlin, U., Farnegardh, M., Logan, D. T., and Nordlund, P. (2001). Structure and function of the radical enzyme ribonucleotide reductase. *Prog. Biophys. Mol. Biol.* 77, 177–268. doi: 10.1016/S0079-6107(01)00014-1
- Eliasson, R., Pontis, E., Fontecave, M., Gerez, C., Harder, J., Jörnvall, H., et al. (1992). Characterization of components of the anaerobic ribonucleotide reductase system from *Escherichia coli*. *J. Biol. Chem.* 267, 25541–25547.
- Eliasson, R., Pontis, E., Jordan, A., and Reichard, P. (1999). Allosteric control of three B12-dependent (class II) ribonucleotide reductases. Implications for the evolution of ribonucleotide reduction. *J. Biol. Chem.* 274, 7182–7189. doi: 10.1074/jbc.274.11.7182
- Eriksson, M., Uhlin, U., Ramaswamy, S., Ekberg, M., Regnström, K., and Sjöberg, B. M. (1997). Binding of allosteric effectors to ribonucleotide reductase protein R1: reduction of active-site cysteines promotes substrate binding. *Structure* 5, 1077–1092. doi: 10.1016/S0969-2126(97)00259-1
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., and Shen, M. Y. (2007). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 15:5.6.1–5.6.30. doi: 10.1002/0471140864.ps0209s50
- Gao, B., and Gupta, R. S. (2012). Phylogenetic framework and molecular signatures for the main clades of the phylum *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* 76, 66–112. doi: 10.1128/MMBR.05011-11
- Garriga, X., Eliasson, R., Torrents, E., Jordan, A., Gibert, I., and Reichard, P. (1996). nrdD and nrdG genes are essential for strict anaerobic growth of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 229, 189–192. doi: 10.1006/bbrc.1996.1778
- Geszvain, K., Gruber, T. M., Mooney, R. A., Gross, C. A., and Landick, R. (2004). A hydrophobic patch on the flap-tip helix of *E. coli* RNA polymerase mediates sigma(70) region 4 function. *J. Mol. Biol.* 343, 569–587. doi: 10.1016/j.jmb.2004.08.063
- Gibson, G. R., Beatty, E. R., Wang, X., and Cummings, J. H. (1995). Selective stimulation of bifidobacteria in the human colon by oligofructose and inulin. *Gastroenterology* 108, 975–982. doi: 10.1016/0016-5085(95)90192-2
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., et al. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 38, W695–W699. doi: 10.1093/nar/gkq313
- Gupta, R. S. (2014). “Chapter 8 - Identification of conserved indels that are useful for classification and evolutionary studies,” in *Methods in Microbiology: New Approaches to Prokaryotic Systematics*, eds M. Goodfellow, I. Sutcliffe, and J. Chun (Cambridge, MA: Academic Press), 153–182.
- Gupta, R. S. (2016). Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol. Rev.* 40, 520–553. doi: 10.1093/femsre/fuw011
- Gupta, R. S., Nanda, A., and Khadka, B. (2017). Novel molecular, structural and evolutionary characteristics of the phosphoketolases from bifidobacteria and *Coriobacteriales*. *PLoS ONE* 12:e0172176. doi: 10.1371/journal.pone.0172176
- Hashimoto, K., and Panchenko, A. R. (2010). Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20352–20357. doi: 10.1073/pnas.1012999107
- Heo, L., Park, H., and Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, W384–W388. doi: 10.1093/nar/gkt458
- Itzhaki, Z., Akiva, E., Altuvia, Y., and Margalit, H. (2006). Evolutionary conservation of domain-domain interactions. *Genome Biol.* 7:R125. doi: 10.1186/gb-2006-7-12-r125
- Jiang, W., Yun, D., Saleh, L., Barr, E. W., Xing, G., Hoffart, L. M., et al. (2007). A manganese(IV)/iron(III) cofactor in *Chlamydia trachomatis* ribonucleotide reductase. *Science* 316, 1188–1191. doi: 10.1126/science.1141179
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Jordan, A., Aslund, F., Pontis, E., Reichard, P., and Holmgren, A. (1997). Characterization of *Escherichia coli* NrdH. A glutaredoxin-like protein with a thioredoxin-like activity profile. *J. Biol. Chem.* 272, 18044–18050. doi: 10.1074/jbc.272.29.18044
- Jordan, A., Pontis, E., Aslund, F., Hellman, U., Gibert, I., and Reichard, P. (1996). The ribonucleotide reductase system of *Lactococcus lactis*. Characterization of an NrdEF enzyme and a new electron transport protein. *J. Biol. Chem.* 271, 8779–8785. doi: 10.1074/jbc.271.15.8779
- Kenyon, C. R., and Osbak, K. (2014). Recent progress in understanding the epidemiology of bacterial vaginosis. *Curr. Opin. Obstet. Gynecol.* 26, 448–454. doi: 10.1097/GCO.0000000000000112
- Khadka, B., and Gupta, R. S. (2017). Identification of a conserved 8 aa insert in the PIP5K protein in the *Saccharomycetaceae* family of fungi and the molecular dynamics simulations and structural analysis to investigate its potential functional role. *Proteins* doi: 10.1002/prot.25306 [Epub ahead of print].
- Kozakov, D., Brenke, R., Comeau, S. R., and Vajda, S. (2006). PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65, 392–406. doi: 10.1002/prot.21117
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Larsson, K. M., Andersson, J., Sjöberg, B. M., Nordlund, P., and Logan, D. T. (2001). Structural basis for allosteric substrate specificity regulation in anaerobic ribonucleotide reductases. *Structure* 9, 739–750. doi: 10.1016/S0969-2126(01)00627-X

- Larsson, K. M., Logan, D. T., and Nordlund, P. (2010). Structural basis for adenosylcobalamin activation in AdoCbl-dependent ribonucleotide reductases. *ACS Chem. Biol.* 5, 933–942. doi: 10.1021/cb1000845
- Leahy, S. C., Higgins, D. G., Fitzgerald, G. F., and van Sinderen, D. (2005). Getting better with bifidobacteria. *J. Appl. Microbiol.* 98, 1303–1315. doi: 10.1111/j.1365-2672.2005.02600.x
- Lee, G. R., Heo, L., and Seok, C. (2016). Effective protein model structure refinement by loop modeling and overall relaxation. *Proteins* 84(Suppl. 1), 293–301. doi: 10.1002/prot.24858
- Logan, D. T., Andersson, J., Sjöberg, B. M., and Nordlund, P. (1999). A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* 283, 1499–1504. doi: 10.1126/science.283.5407.1499
- Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., et al. (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50, 437–450. doi: 10.1002/prot.10286
- Lundin, D., Gribaldo, S., Torrents, E., Sjöberg, B. M., and Poole, A. M. (2010). Ribonucleotide reduction - horizontal transfer of a required function spans all three domains. *BMC Evol. Biol.* 10:383. doi: 10.1186/1471-2148-10-383
- Lundin, D., Torrents, E., Poole, A. M., and Sjöberg, B. M. (2009). RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. *BMC Genomics* 10:589. doi: 10.1186/1471-2164-10-589
- Luthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* 356, 83–85. doi: 10.1038/356083a0
- Lyskov, S., Chou, F. C., Conchuir, S. O., Der, B. S., Drew, K., Kuroda, D., et al. (2013). Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PLoS ONE* 8:e63906. doi: 10.1371/journal.pone.0063906
- Lyskov, S., and Gray, J. J. (2008). The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* 36, W233–W238. doi: 10.1093/nar/gkn216
- Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–W331. doi: 10.1093/nar/gkh454
- Masco, L., Huys, G., De Brandt, E., Temmerman, R., and Swings, J. (2005). Culture-dependent and culture-independent qualitative analysis of probiotic products claimed to contain bifidobacteria. *Int. J. Food Microbiol.* 102, 221–230. doi: 10.1016/j.ijfoodmicro.2004.11.018
- Milani, C., Lugli, G. A., Duranti, S., Turrioni, F., Bottacini, F., Mangifesta, M., et al. (2014). Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Appl. Environ. Microbiol.* 80, 6290–6302. doi: 10.1128/AEM.02308-14
- Milani, C., Turrioni, F., Duranti, S., Lugli, G. A., Mancabelli, L., Ferrario, C., et al. (2015). Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment. *Appl. Environ. Microbiol.* 82, 980–991. doi: 10.1128/AEM.03500-15
- Mills, S., Stanton, C., Fitzgerald, G. F., and Ross, R. P. (2011). Enhancing the stress responses of probiotics for a lifestyle from gut to product and back again. *Microb. Cell Fact.* 10(Suppl. 1):S19. doi: 10.1186/1475-2859-10-S1-S19
- Nordlund, P., and Reichard, P. (2006). Ribonucleotide reductases. *Annu. Rev. Biochem.* 75, 681–706. doi: 10.1146/annurev.biochem.75.103004.142443
- Oberg, T. S., Steele, J. L., Ingham, S. C., Smeianov, V. V., Briczinski, E. P., Abdalla, A., et al. (2011). Intrinsic and inducible resistance to hydrogen peroxide in *Bifidobacterium* species. *J. Ind. Microbiol. Biotechnol.* 38, 1947–1953. doi: 10.1007/s10295-011-0983-y
- Ollagnier, S., Mulliez, E., Gaillard, J., Eliasson, R., Fontecave, M., and Reichard, P. (1996). The anaerobic *Escherichia coli* ribonucleotide reductase. Subunit structure and iron sulfur center. *J. Biol. Chem.* 271, 9410–9416. doi: 10.1074/jbc.271.16.9410
- Palframan, R. J., Gibson, G. R., and Rastall, R. A. (2003). Carbohydrate preferences of *Bifidobacterium* species isolated from the human gut. *Curr. Issues Intest. Microbiol.* 4, 71–75.
- Petersson, L., Graslund, A., Ehrenberg, A., Sjöberg, B. M., and Reichard, P. (1980). The iron center in ribonucleotide reductase from *Escherichia coli*. *J. Biol. Chem.* 255, 6706–6712.
- Poole, A. M., Logan, D. T., and Sjöberg, B. M. (2002). The evolution of the ribonucleotide reductases: much ado about oxygen. *J. Mol. Evol.* 55, 180–196. doi: 10.1007/s00239-002-2315-3
- Reichard, P. (1993). From RNA to DNA, why so many ribonucleotide reductases? *Science* 260, 1773–1777.
- Reichard, P. (2010). Ribonucleotide reductases: substrate specificity by allostery. *Biochem. Biophys. Res. Commun.* 396, 19–23. doi: 10.1016/j.bbrc.2010.02.108
- Roca, I., Torrents, E., Sahlin, M., Gibert, I., and Sjöberg, B. M. (2008). NrdI essentiality for class Ib ribonucleotide reduction in *Streptococcus pyogenes*. *J. Bacteriol.* 190, 4849–4858. doi: 10.1128/JB.00185-08
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. doi: 10.1038/nprot.2010.5
- Sanders, M. E. (1998). Development of consumer probiotics for the US market. *Br. J. Nutr.* 80, S213–S218.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367. doi: 10.1093/nar/gki481
- Schoeffler, A. J., May, A. P., and Berger, J. M. (2010). A domain insertion in *Escherichia coli* GyrB adopts a novel fold that plays a critical role in gyrase function. *Nucleic Acids Res.* 38, 7830–7844. doi: 10.1093/nar/gkq665
- Schrödinger, L. (2016). *The PyMOL Molecular Graphics System, Version 1.8*. Available at: <https://www.pymol.org/>
- Shen, M. Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507–2524. doi: 10.1110/ps.062416606
- Singh, B., and Gupta, R. S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genomics* 281, 361–373. doi: 10.1007/s00438-008-0417-3
- Sintchak, M. D., Arjara, G., Kellogg, B. A., Stubbe, J., and Drennan, C. L. (2002). The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. *Nat. Struct. Biol.* 9, 293–300. doi: 10.1038/nsb774
- Sippl, M. J., Lackner, P., Domingues, F. S., and Koppensteiner, W. A. (1999). An attempt to analyse progress in fold recognition from CASP1 to CASP3. *Proteins Suppl.* 3, 226–230. doi: 10.1002/(SICI)1097-0134(1999)37:3+<226::AID-PROT29>3.0.CO;2-Z
- Smith, S. M., Ogbara, T., and Eng, R. H. (1992). Involvement of *Gardnerella vaginalis* in urinary tract infections in men. *J. Clin. Microbiol.* 30, 1575–1577.
- Sun, X., Eliasson, R., Pontis, E., Andersson, J., Buist, G., Sjöberg, B. M., et al. (1995). Generation of the glycyl radical of the anaerobic *Escherichia coli* ribonucleotide reductase requires a specific activating enzyme. *J. Biol. Chem.* 270, 2443–2446. doi: 10.1074/jbc.270.6.2443
- Tamao, Y., and Blakley, R. L. (1973). Direct spectrophotometric observation of an intermediate formed from deoxyadenosylcobalamin in ribonucleotide reduction. *Biochemistry* 12, 24–34. doi: 10.1021/bi00725a005
- Tamura, K., Stecher, G., Peterson, D., Filipinski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Thelander, L., and Reichard, P. (1979). Reduction of ribonucleotides. *Annu. Rev. Biochem.* 48, 133–158. doi: 10.1146/annurev.bi.48.070179.001025
- Torrents, E. (2014). Ribonucleotide reductases: essential enzymes for bacterial life. *Front. Cell Infect. Microbiol.* 4:52. doi: 10.3389/fcimb.2014.00052
- Torrents, E., Aloy, P., Gibert, I., and Rodriguez-Trelles, F. (2002). Ribonucleotide reductases: divergent evolution of an ancient enzyme. *J. Mol. Evol.* 55, 138–152. doi: 10.1007/s00239-002-2311-7
- Torrents, E., Eliasson, R., Wolpher, H., Graslund, A., and Reichard, P. (2001). The anaerobic ribonucleotide reductase from *Lactococcus lactis*. Interactions between the two proteins NrdD and NrdG. *J. Biol. Chem.* 276, 33488–33494. doi: 10.1074/jbc.M103743200
- Torrents, E., Jordan, A., Karlsson, M., and Gibert, I. (2000). Occurrence of multiple ribonucleotide reductase classes in gamma-proteobacteria species. *Curr. Microbiol.* 41, 346–351. doi: 10.1007/s002840010147
- Turrioni, F., Ribbera, A., Foroni, E., van Sinderen, D., and Ventura, M. (2008). Human gut microbiota and bifidobacteria: from composition to functionality. *Antonie Van Leeuwenhoek* 94, 35–50. doi: 10.1007/s10482-008-9232-4
- Turrioni, F., van Sinderen, D., and Ventura, M. (2009). Bifidobacteria: from ecology to genomics. *Front. Biosci.* 14:4673–4684. doi: 10.2741/3559
- Turrioni, F., Ventura, M., Butto, L. F., Duranti, S., O'Toole, P. W., Motherway, M. O., et al. (2014). Molecular dialogue between the human gut microbiota and the host: a *Lactobacillus* and *Bifidobacterium* perspective. *Cell Mol. Life Sci.* 71, 183–203. doi: 10.1007/s00018-013-1318-0

- Uhlin, U., and Eklund, H. (1994). Structure of ribonucleotide reductase protein R1. *Nature* 370, 533–539. doi: 10.1038/370533a0
- Uppsten, M., Farnegardh, M., Domkin, V., and Uhlin, U. (2006). The first holocomplex structure of ribonucleotide reductase gives new insight into its mechanism of action. *J. Mol. Biol.* 359, 365–377. doi: 10.1016/j.jmb.2006.03.035
- Uppsten, M., Farnegardh, M., Jordan, A., Eliasson, R., Eklund, H., and Uhlin, U. (2003). Structure of the large subunit of class Ib ribonucleotide reductase from *Salmonella typhimurium* and its complexes with allosteric effectors. *J. Mol. Biol.* 330, 87–97. doi: 10.1016/S0022-2836(03)00538-2
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., et al. (2007). Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.* 71, 495–548. doi: 10.1128/MMBR.00005-07
- Ventura, M., O'Flaherty, S., Claesson, M. J., Turrone, F., Klaenhammer, T. R., van Sinderen, D., et al. (2009). Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat. Rev. Microbiol.* 7, 61–71. doi: 10.1038/nrmicro2047
- Ventura, M., Turrone, F., Lugli, G. A., and van Sinderen, D. (2014). Bifidobacteria and humans: our special friends, from ecological to genomics perspectives. *J. Sci. Food Agric.* 94, 163–168. doi: 10.1002/jsfa.6356
- Wei, Y., Thompson, J., and Floudas, C. A. (2011). CONCORD: a consensus method for protein secondary structure prediction via mixed integer linear optimization. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 468, 831–850. doi: 10.1098/rspa.2011.0514
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851
- Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. doi: 10.1093/nar/gkm290
- Xiao, M., Xu, P., Zhao, J., Wang, Z., Zuo, F., Zhang, J., et al. (2011). Oxidative stress-related responses of *Bifidobacterium longum* subsp. *longum* BBMN68 at the proteomic level after exposure to oxygen. *Microbiology* 157, 1573–1588. doi: 10.1099/mic.0.044297-0
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12, 7–8. doi: 10.1038/nmeth.3213
- Zhang, G., Gao, B., Adeolu, M., Khadka, B., and Gupta, R. S. (2016). Phylogenomic analyses and comparative studies on genomes of the *Bifidobacteriales*: identification of molecular signatures specific for the order *Bifidobacteriales* and its different subclades. *Front. Microbiol.* 7:978. doi: 10.3389/fmicb.2016.00978
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40. doi: 10.1186/1471-2105-9-40
- Zhi, X. Y., Li, W. J., and Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int. J. Syst. Evol. Microbiol.* 59, 589–608. doi: 10.1099/ijs.0.65780-0
- Zuo, F., and Chen, S. (2016). Recently performed Next generation sequencing based transcriptomic analysis of oxidative stress response in *Bifidobacterium longum* BBMN68 (data accessible at NCBI GEO database (Edgar et al., 2002), accession GSE65320). NCBI GEO database. Available at: <http://www.ncbi.nlm.nih.gov/geo/>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PL and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Alnajar, Khadka and Gupta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Characterization of the Sorbitol Utilization Cluster of the Probiotic *Pediococcus parvulus* 2.6: Genetic, Functional and Complementation Studies in Heterologous Hosts

Adrian Pérez-Ramos<sup>1</sup>, Maria L. Werning<sup>1,2</sup>, Alicia Prieto<sup>1</sup>, Pasquale Russo<sup>3</sup>, Giuseppe Spano<sup>3</sup>, Mari L. Mohedano<sup>1</sup> and Paloma López<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Maria Jesus Yebra,  
Instituto de Agroquímica y Tecnología  
de Alimentos (CSIC), Spain  
Bopda Waffo Alain,  
Alabama State University,  
United States  
Antonius Suwanto,  
Bogor Agricultural University,  
Indonesia

### \*Correspondence:

Paloma López  
plg@cib.csic.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 28 September 2017

**Accepted:** 20 November 2017

**Published:** 05 December 2017

### Citation:

Pérez-Ramos A, Werning ML,  
Prieto A, Russo P, Spano G,  
Mohedano ML and López P (2017)  
Characterization of the Sorbitol  
Utilization Cluster of the Probiotic  
*Pediococcus parvulus* 2.6: Genetic,  
Functional and Complementation  
Studies in Heterologous Hosts.  
Front. Microbiol. 8:2393.  
doi: 10.3389/fmicb.2017.02393

<sup>1</sup> Biological Research Center (CIB), Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Center of Research and Transfer of Catamarca (CITCA), Consejo Nacional de Investigaciones Científicas y Técnicas, Catamarca, Argentina,

<sup>3</sup> Department of Agricultural, Food and Environmental Sciences, University of Foggia, Foggia, Italy

*Pediococcus parvulus* 2.6 secretes a 2-substituted (1,3)- $\beta$ -D-glucan with prebiotic and immunomodulatory properties. It is synthesized by the GTF glycosyltransferase using UDP-glucose as substrate. Analysis of the *P. parvulus* 2.6 draft genome revealed the existence of a sorbitol utilization cluster of six genes (*gutFRMCBA*), whose products should be involved in sorbitol utilization and could generate substrates for UDP-glucose synthesis. Southern blot hybridization analysis showed that the cluster is located in a plasmid. Analysis of metabolic fluxes and production of the exopolysaccharide revealed that: (i) *P. parvulus* 2.6 is able to metabolize sorbitol, (ii) sorbitol utilization is repressed in the presence of glucose and (iii) sorbitol supports the synthesis of 2-substituted (1,3)- $\beta$ -D-glucan. The sorbitol cluster encodes two putative regulators, GutR and GutM, in addition to a phosphoenolpyruvate-dependent phosphotransferase transport system and sorbitol-6-phosphate dehydrogenase. Therefore, we investigated the involvement of GutR and GutM in the expression of *gutFRMCBA*. The promoter-probe vector pPCR based on the *mrfp* gene, which encodes the fluorescence protein mCherry, was used to test the potential promoter of the cluster ( $P_{gut}$ ) and the genes encoding the regulators. This was performed by transferring by electrotransformation the recombinant plasmids into two hosts, which metabolize sorbitol: *Lactobacillus plantarum* and *Lactobacillus casei*. Upon growth in the presence of sorbitol, but not of glucose, only the presence of  $P_{gut}$  was required to support expression of *mrfp* in *L. plantarum*. In *L. casei* the presence of sorbitol in the growth medium and the pediococcal *gutR* or *gutR* plus *gutM* in the genome was required for  $P_{gut}$  functionality. This demonstrates that: (i)  $P_{gut}$  is required for expression of the *gut* cluster, (ii)  $P_{gut}$  is subjected to catabolic repression in lactobacilli, (iii) GutR is an activator, and (iv) in the presence of sorbitol, *trans*-complementation for activation of  $P_{gut}$  exists in *L. plantarum* but not in *L. casei*.

**Keywords:** *Pediococcus parvulus*, exopolysaccharides,  $\beta$ -glucans, sorbitol, lactic acid bacteria, probiotic

## INTRODUCTION

Sorbitol, also named D-glucitol, is a six-carbon sugar polyol widespread in plants, particularly in fruits, such as berries, cherries, plums, pears and apples. However, sorbitol is obtained industrially, by catalytic hydrogenation of glucose or glucose/fructose mixtures. This polyol has a relative sweetness of about 60% compared to that of sucrose, high-water solubility and is largely used as a low calorie sweetener, humectant, texturizer and softener (Zumbé et al., 2001). In addition, sorbitol is used in the production of pharmaceutical compounds, such as sorbose and ascorbic acid, and as a vehicle for drug-suspension (Silveira and Jonas, 2002). Sorbitol has also a potential prebiotic effect *in vivo*, since it does not contribute to the formation of dental caries, is slowly and only partially absorbed in the small intestine and can reach the colon where it can act as substrate for bacterial fermentation. Supplementation with sorbitol resulted in enrichment of lactobacilli in rat colon and cecum (Sarmiento-Rubiano et al., 2007).

Sorbitol absorption is mediated by dose and concentration. Doses greater than 30 g can cause water retention, resulting in osmotic diarrhea, bloating, flatulence, cramping and abdominal pain (Fernández-Bañares et al., 2009). These doses vary depending on the condition of the intestinal absorption surface. In patients with malabsorption, the ingestion of 5–20 g, provoked diarrhea and gastrointestinal complications (Montalto et al., 2013). In the colon, this sugar alcohol is metabolized by some species of *Lactobacillus* and is also a preferred carbon source for human intestinal bifidobacteria (Sarmiento-Rubiano et al., 2007).

Furthermore, utilization of sorbitol as a carbon source has been described in a variety of bacteria within the filo proteobacteria (Yamada and Saier, 1988; Aldridge et al., 1997) and firmicutes (Tangney et al., 1998; Boyd et al., 2000; Yebra and Pérez-Martínez, 2002). Among the firmicutes, there are some lactic acid bacteria (LAB) with catabolic pathways for sorbitol metabolism (Rhodes and Kator, 1999; Sarmiento-Rubiano et al., 2007). These pathways are encoded by genes organized in *gut* operons, and include the sorbitol transport system, sorbitol-6-phosphate dehydrogenase (S6PD) as well as regulatory protein(s), and those of *Lactobacillus casei* and *Lactobacillus plantarum* have been characterized (Nissen et al., 2005; Ladero et al., 2007; Alcantara et al., 2008).

Sorbitol is transported into the cells and phosphorylated to sorbitol-6-phosphate by a phosphopyruvate-dependent phosphotransferase (PTS) sorbitol system (PTS<sup>sut</sup>). Each PTS is composed of two cytoplasmic enzymes, common to the transport of different compounds (EI and HPr) and of different membrane-associated enzyme complexes (EII), specific for one, or several substrates. The genes *gutC*, *gutB* and *gutA* encode the EII domain of a sorbitol PTS (Alcantara et al., 2008). The *gutF* gene encodes a sorbitol-6-P dehydrogenase, which catalyzes the conversion of sorbitol-6-phosphate to fructose-6-phosphate, a compound that is introduced into the glycolytic pathway with NADH regeneration (Nissen et al., 2005). The *gutR* and *gutM* genes encode two regulatory proteins. The role of the GutM and GutR proteins has been studied in *Escherichia coli*, operating GutM as an activator and GutR as a repressor (Yamada and

Saier, 1988). In the firmicutes group, the analyzed *gut* operons contain homologs to the *gutM* and *gutR* genes, but the role of GutR regulator is different from that of *E. coli*. The GutR of *L. casei* has been functionally characterized and it has been shown to be a PTS-controlled transcriptional activator, via a PTS regulation binding domain (PRD) (Stülke et al., 1998). Also, both the GutR binding sequence and the PRD domain are conserved in firmicutes. GutM encodes a highly conserved protein in firmicutes and in *L. casei* plays a regulatory role (Alcantara et al., 2008).

*Pediococcus parvulus* 2.6 (Werning et al., 2006) (previously named *Pediococcus damnosus*) is a lactic acid bacteria isolated from a rosy cider (Fernández et al., 1995). This LAB produces a 2-substituted (1,3)-β-D-glucan exopolysaccharide (EPS) (Dueñas-Chasco et al., 1997), with high molecular mass (>10<sup>6</sup> Da), and whose rheological properties showed its potential utility as a biothickening agent (Velasco et al., 2009). The presence of this EPS improves some probiotic features of *P. parvulus* 2.6, including tolerance to simulated gastrointestinal conditions and adherence to Caco-2 cell lines and reduces inflammation-related cytokine levels produced by polarized macrophages (Fernández de Palencia et al., 2009; Immerstrand et al., 2010). Moreover, the purified EPS improves the growth, viability and adhesion capability of probiotic microorganisms (Russo et al., 2012), also it activates macrophages with anti-inflammatory effects (Notararigo et al., 2014), and decreases the levels of the proinflammatory IL8 in human intestine cultures (Notararigo et al., unpublished data). The draft genome of *P. parvulus* 2.6 has been determined (Pérez-Ramos et al., 2016), and its analysis showed the existence of a putative sorbitol utilization *gut* operon in this bacterium. Thus, this current work focuses on the genomic location, expression and metabolic involvement of the *gut* operon of *P. parvulus* 2.6 in sorbitol catabolism, as well as its interplay with EPS production by this bacterium.

## MATERIALS AND METHODS

### Bacterial Strains and Growth Conditions

The bacteria used in this work are listed in Table 1. *Pediococcus* and *Lactobacillus* strains were routinely grown in de Man Rogosa Sharpe (MRS) broth (Pronadisa, Madrid, Spain) at 30°C and 37°C, respectively. *Lactococcus lactis* strains were grown in ESTY broth (Pronadisa) supplemented with 0.5% glucose at 30°C. When bacteria carried the pPCR plasmid or its derivatives the medium was supplemented with chloramphenicol (Cm) at 5 μg mL<sup>-1</sup> for *L. lactis* and at 10 μg mL<sup>-1</sup> for lactobacilli. *E. coli* V517 was grown in LB broth and incubated at 37°C.

For evaluation of sorbitol utilization, *P. parvulus* strains were grown in a MRS broth made by components (de Man et al., 1960) without glucose, pH was adjusted to 5.2 and the medium supplemented with 10 mM glucose (MRSG), 30 mM sorbitol (MRSS) or 10 mM glucose plus 30 mM sorbitol (MRSGS) at 30°C. Prior selection of conditions for growth in presence of sorbitol several tests were performed. First various carbon sources were tested (10 mM glucose, 10 mM fructose or 10 mM maltose) and

**TABLE 1 |** Bacteria used in this work.

Bacteria	Plasmid	Resistance	Characteristics	Reference
<i>Pediococcus parvulus</i> 2.6	pPP1, pPP2, and pPP3	–	2-substituted (1,3)- $\beta$ -D-glucan producer	Pérez-Ramos et al., 2016
<i>P. parvulus</i> 2.6NR	pPP1 and pPP3	–	Non-EPS-producing strain. Derivative of 2.6 strain by pPP2 plasmid curing	Fernández et al., 1995
<i>Lactococcus lactis</i> subsp. <i>cremoris</i> MG1363	–	–	Plasmid free type strain used for plasmid cloning	Wegmann et al., 2007
<i>L. lactis</i> subsp. <i>cremoris</i> MG1363[pRCR]	pRCR	Cm <sup>R</sup>	Source of promoter probe pRCR containing the <i>mrfp</i> gene, which encodes the fluorescent mCherry protein	Mohedano et al., 2015
<i>Escherichia coli</i> V517	8 plasmids pVA517A through pVA517H	ND	Source of plasmids used as references in agarose gel analysis	Macrina et al., 1978
<i>Lactobacillus casei</i> BL23	–	–	Bacteria used for heterologous gene expression	Mazé et al., 2010
<i>L. casei</i> BL23[pRCR16]	[pRCR16]	Cm <sup>R</sup>	Derivative of pRCR by cloning of <i>P<sub>gut</sub></i> upstream of <i>mrfp</i>	This study
<i>L. casei</i> BL23[pRCR17]	[pRCR17]	Cm <sup>R</sup>	Derivative of pRCR16 by cloning of <i>gutR</i> upstream of <i>mrfp</i>	This study
<i>L. casei</i> BL23[pRCR18]	[pRCR18]	Cm <sup>R</sup>	Derivative of pRCR16 by cloning of <i>gutM</i> upstream of <i>mrfp</i>	This study
<i>L. casei</i> BL23[pRCR19]	[pRCR19]	Cm <sup>R</sup>	Derivative of pRCR16 by cloning of <i>gutMR</i> upstream of <i>mrfp</i>	This study
<i>Lactobacillus plantarum</i> 90	1 uncharacterized plasmid	ND	Bacteria used for heterologous gene expression	Lamontanara et al., 2015
<i>L. plantarum</i> 90[pRCR16]	[pRCR16]	Cm <sup>R</sup>	Derivative of pRCR by cloning of <i>P<sub>gut</sub></i> upstream of <i>mrfp</i>	This study
<i>L. plantarum</i> 90[pRCR17]	[pRCR17]	Cm <sup>R</sup>	Derivative of pRCR16 by cloning of <i>gutR</i> upstream of <i>mrfp</i>	This study
<i>L. plantarum</i> 90[pRCR18]	[pRCR18]	Cm <sup>R</sup>	Derivative of pRCR16 by cloning of <i>gutM</i> upstream of <i>mrfp</i>	This study
<i>L. plantarum</i> 90[pRCR19]	[pRCR19]	Cm <sup>R</sup>	Derivative of pRCR16 by cloning of <i>gutMR</i> upstream of <i>mrfp</i>	This study

ND, no determined; Cm<sup>R</sup>, resistance to chloramphenicol.

pH at 6.8, 5.2 or 4.0 and then influence of aeration was evaluated in presence of 10 mM glucose at either pH 6.8 and 5.2 (results not show).

For evaluation of mCherry expression, *Lactobacillus* strains were grown in a MRSG containing 55 mM (1% w/v) glucose or in a MRSS containing 55 mM (1% w/v) sorbitol at 37°C.

## Plasmidic DNA Preparations

Total plasmidic DNA preparations of *P. parvulus* 2.6 and 2.6NR strains were prepared as follows. Bacterial cultures were grown to an optical density at 600 nm (OD<sub>600 nm</sub>) of 2.5, and 100 mL of each culture were sedimented by centrifugation at 10,000  $\times$  g for 20 min at 4°C. The cells were resuspended in 4 mL of a solution containing 50 mM Tris/HCl pH 8.0, 10 mM EDTA, lysozyme (30 mg mL<sup>-1</sup>) and RNase A (10  $\mu$ g mL<sup>-1</sup>), and incubated for 30 min at 37°C. Then, 4 mL of a solution containing 220 mM NaOH and 1.33% sodium dodecyl sulfate) were added and samples were incubated for 5 min at room temperature. Upon addition of 5 M potassium acetate pH 5.0 (4 mL), samples were centrifugated at 10,000  $\times$  g for 15 min at 21°C. The DNA present in the supernatants was precipitated, concentrated by addition of 8.7 mL of isopropanol, sedimented by centrifugation at 10,000  $\times$  g for 15 min at 4°C, and resuspended in 10 mM Tris, 1 mM EDTA buffer (4.3 mL).

The DNA preparation was deproteinated by treatment with 7.5 M ammonium acetate (2.7 mL) and phenol (4.3 mL) during 5 min at room temperature and then sedimented at 10,000  $\times$  g for 5 min at 21°C. The aqueous phase containing total plasmidic DNA was further purified by isopycnic CsCl density gradient centrifugation and dialysis as previously described (López et al., 1989). The final recovery was 54  $\mu$ g and 58  $\mu$ g for 2.6 and 2.6NR DNA preparations, respectively.

The recombinant plasmids from the lactococcal and lactobacilli strains were isolated using the High pure plasmid isolation kit (Roche) as follows. Bacteria were grown until stationary phase (10<sup>9</sup> colony forming units mL<sup>-1</sup>) and 1 mL of each culture were sedimented by centrifugation at 10,000  $\times$  g for 10 min at 4°C. Cells were resuspended in solution I of the kit supplemented with lysozyme (30 mg mL<sup>-1</sup>) and were incubated for 30 min at 37°C. Then, plasmid isolation were performed as described in the kit protocol, eluting the plasmidic DNA in 100  $\mu$ L at approximately 100 ng  $\mu$ L<sup>-1</sup>.

## Sequencing

DNA sequencing was performed by the dideoxy method at Secugen (Madrid, Spain). The sequencing of the sorbitol utilization cluster and the flanking regions of pPP1 of *P. parvulus* 2.6 was performed using total plasmidic DNA preparations

of the bacterium (see above) with the walking strategy and the sequence has been deposited in GenBank (accession No MF766019). The lack of sorbitol cluster in the 2.6NR strain was confirmed by sequencing of its pPP1 plasmid by using as substrates a total plasmidic preparation of 2.6NR strain and either pPP1\*F or pPP1\*R primers (see **Table 2**). In addition, in the case of sequencing with pPP1\*F, it was also used as substrate the product of a polymerization reaction catalyzed by the bacteriophage  $\Phi$ 29 DNA polymerase with plasmidic DNA of *P. parvulus* 2.6NR and hexamers containing random sequences.

## Construction of pRCR16, pRCR17, pRCR18, and pRCR19

A region located upstream of the *P. parvulus* 2.6 *gut* operon carrying the putative  $P_{gut}$  promoter and the *gutR* and *gutM* genes was cloned into the promoter probe pRCR vector. To this end, three DNA regions of pPP1 plasmid were amplified with Phusion High Fidelity Polymerase (PHFP, ThermoFisher Scientific) by using a plasmidic DNA preparation of *P. parvulus* 2.6 and the primers depicted in **Table 2**, which have homology with pPP1 DNA and carry restriction sites suitable for cloning. Plasmid pRCR16 (**Figure 1**) was generated by ligation of the  $P_{gut}$  promoter to the pRCR promoter probe vector (Mohedano et al., 2015), after double digestion of both DNAs with BglII and XmaI (New England Biolabs, Ipswich, MA, United States), with the T4 DNA ligase (New England Biolabs). Then, between the XmaI and XbaI restriction sites of pRCR16 three amplicons were independently cloned, containing *gutR*, *gutM* or *gutRM*, generating plasmids pRCR17, pRCR18 and pRCR19, respectively. The clonings were performed in *L. lactis* MG1363, the ligations mixtures were used to transform the bacteria by electroporation (25  $\mu$ F,

2.5 kV and 200  $\Omega$  in 0.2 cm cuvettes), as previously described (Dornan and Collins, 1987) and transformants were selected in ESTY-agar plates supplemented with Cm at 5  $\mu$ g mL<sup>-1</sup>. The inserts present in the new four recombinant plasmids were confirmed by automated sequencing. Then, DNA preparations of pRCR17, pRCR18 and pRCR19 obtained from *L. lactis* MG1363 (0.5  $\mu$ g) were used for transfer to lactobacilli by electroporation (25  $\mu$ F, 1.3 kV and 200  $\Omega$  in 0.1 cm cuvettes) as previously described (Berthier et al., 1996) and transformants were selected in MRS-agar plates supplemented with Cm at 10  $\mu$ g mL<sup>-1</sup>.

## Southern Hybridization

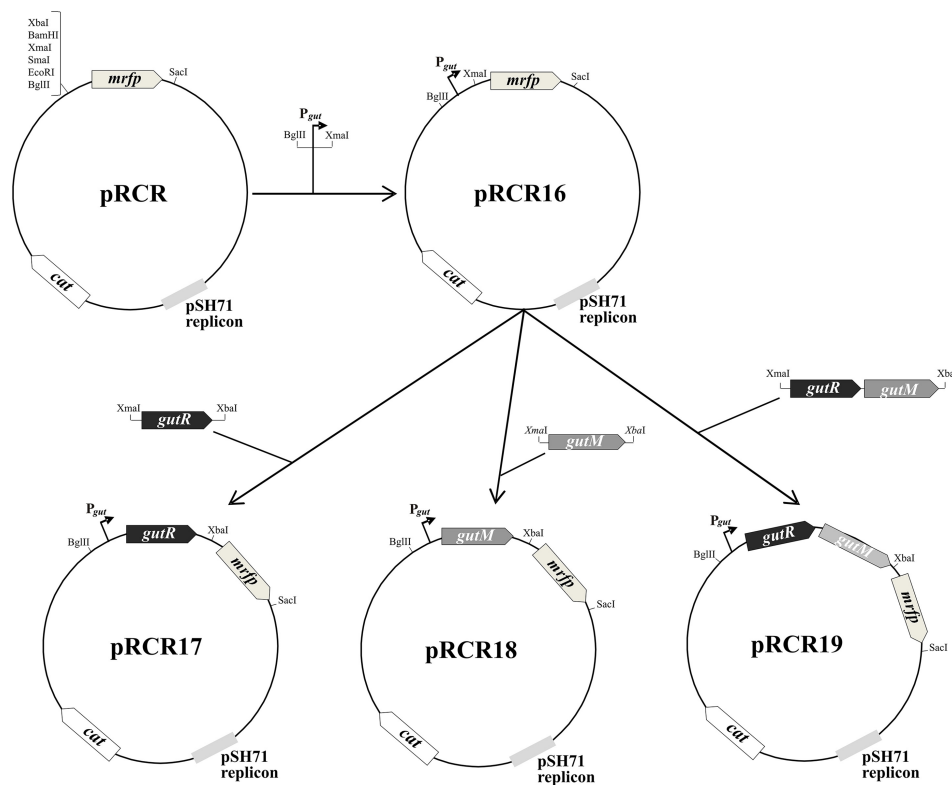
Plasmid samples were fractionated by electrophoresis in a 0.7% agarose gel and DNA molecules were revealed by staining with ethidium bromide at 0.5  $\mu$ g mL<sup>-1</sup>. The image of the gels was obtained with GelDoc 200 (BioRad) and the bands were quantitated with the Quantity One 4.5.2 software (BioRad). The DNA fragments were transferred to a nylon membrane Biodyne A (PALL Gelman Laboratory, Ann Arbor, MI, United States) by 5 inches Hg of vacuum for 2 h using the Vacuum Blotter model 785 (Bio-Rad). Internal regions of *gutF*, *gutR* and *gutB* genes were amplified by PCR generating amplicons 1, 2, and 3, respectively, in reactions catalyzed by PHFP, and by using as substrate total plasmidic DNA preparation of *P. parvulus* 2.6 and the primer pairs shown in **Table 2**. Then, the amplicons were labeled with digoxigenin-dUTP by using the DIG high prime DNA labeling and detection starter kit II (Roche, Mannheim, Germany). Each DIG-labeled DNA probe (25 ng mL<sup>-1</sup>) was used for hybridization at 45°C following the specifications of the kit's supplier. The hybridization bands were revealed with the chemiluminescent substrate CSPD, and the signals were detected

**TABLE 2** | Oligonucleotides used in this work.

Primers	Sequence (5' - 3')	Utilization	Amplicon size (bp)
pts1F	TGCGGAAGCGGTTAATCGGCT	<i>gutF</i> DNA probe <sup>a</sup>	618
pts1R	CCACGACTCTTGCCCTCCGCA		
ptsRF	CGAACTGGAAGCAACCTGGGA		
ptsRR	CCGATGAATAATTGGCGCTGC	<i>gutR</i> DNA probe <sup>a</sup>	652
ptsBF	GGAATGGAAGCTGTTGATGGC		
ptsBR	CAACGCCAATCAAGGTCCCGA		
pgutBgIII	GAAGATCTACCATATGGCGATAATGAAAA	Cloning of $P_{gut}$ in pRCR <sup>a</sup>	186
pgutXmaIR	TCGCTCCCGGGTCATTTCTTTTC	Cloning of <i>gutR</i> in pRCR16 <sup>a</sup>	1952
gutRXmaIF	CGTGGTTAACC CGGGAATTTAGTTG		
gutRXbaIR	GCTCTAGAAACGCACTGACTAGGATCA		
gutMXmaIF	TCCCCCGGGTTAAATCAGTTGATGGA	Cloning of <i>gutM</i> in pRCR16 <sup>a</sup>	597
gutMXbaIR	GCTCTAGAACAGCCCATAAGCCC		
gutRXmaIF	CGTGGTTAACC CGGGAATTTAGTTG		
gutMXbaIR	GCTCTAGAACAGCCCATAAGCCC	Cloning of <i>gutRM</i> in pRCR16 <sup>a</sup>	2472
pPP1*F	CATAGTTCACTGGGCTACCA		
pPP1*R	TAGCGGTGCCTCCCTTTAAT		
		Sequencing of pPP1* <sup>b</sup>	–
			–

<sup>a</sup>Plasmidic DNA preparations of *P. parvulus* 2.6 was used as substrate for the PCR reactions. <sup>b</sup>Plasmidic DNA preparations of *P. parvulus* 2.6NR was used as substrate for the DNA sequencing.





**FIGURE 1 |** Scheme of construction of plasmids pRCR16, pRCR17, pRCR18, and pRCR19. Maps of these plasmids and of the parental pRCR are depicted. The pertinent restriction sites are shown, as well as the putative promoter of the *gut* operon ( $P_{gut}$ ) and the following genes: *mrfp*, *cat*, *gutR* and *gutM* encoding the mCherry, a chloramphenicol acetyltransferase, GutR and GutM, respectively.

with the LAS-3000 imaging system (Fujifilm, Stamford, CT, United States).

## Analysis of the Metabolic Fluxes of *P. parvulus* and Its EPS Production

*P. parvulus* 2.6 and 2.6NR strains were grown in either MRSG or MRSGS under aerobic conditions (shaking at 180 rpm), at 30°C during 66 h, and samples were taken at the times indicated in **Figure 2** to monitor growth by determination of optical density at 600 nm and of acidification of the media by measuring pH. Also, samples were centrifuged at  $16,000 \times g$  for 30 min at 4°C, and the levels of glucose, sorbitol, lactic acid and EPS in the supernatants were analyzed. The experiments were performed in triplicate for each strain and in each condition of growth.

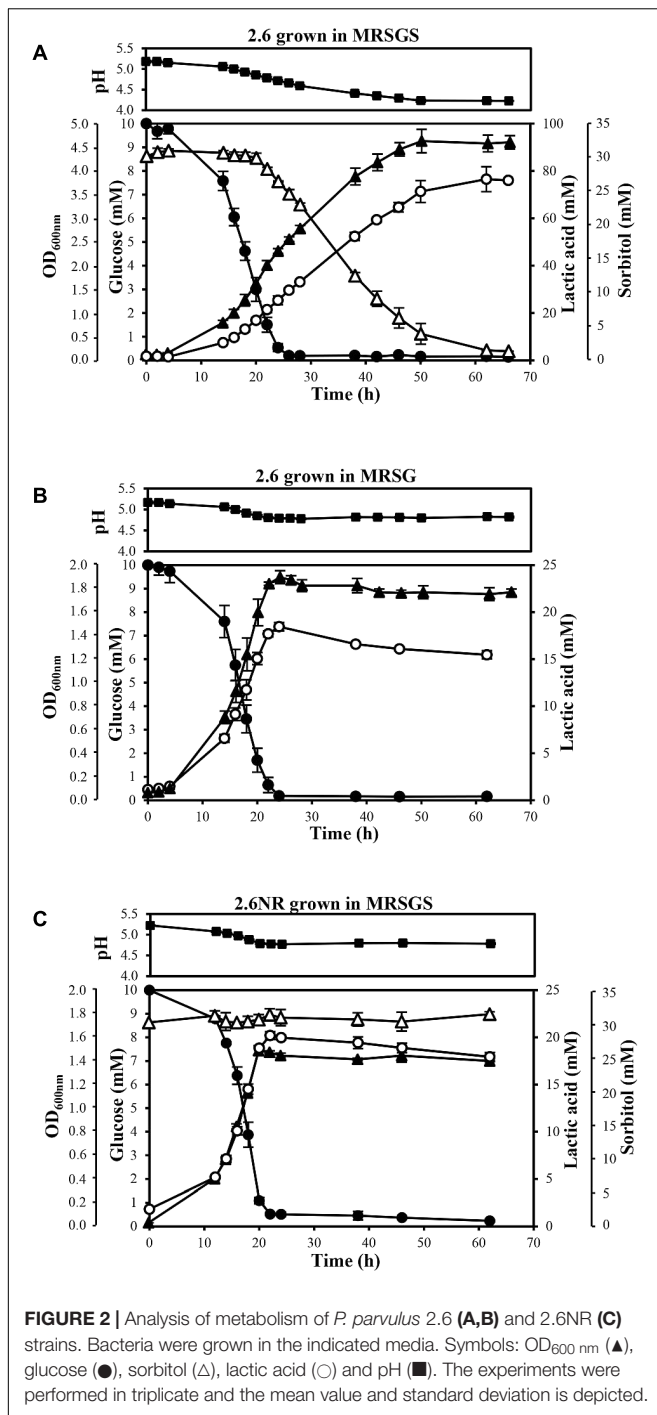
## Analysis of Culture Supernatants by Gas Chromatography-Mass Spectrometry (GC-MS)

The concentration of glucose, sorbitol and lactic acid was determined by GC-MS using myo-inositol as internal standard. For this analysis, myo-inositol (100 µg) was first added to aliquots of the bacterial culture supernatants. The mixture was lyophilized and derivatized with 2.5% hydroxylamine chloride in pyridine for 30 min at 70°C, to form the sugar oximes. Afterward, bis-trimethylsilyl trifluoroacetamide (BSTFA) was added and

samples were incubated for 45 min at 80°C, to form the trimethylsilylated derivatives. Identification and quantification of the compounds were performed by GC-MS on a 7980A-5975C instrument (Agilent, Santa Clara, CA, United States) equipped with a HP-5MS column (30 m  $\times$  0.25 mm I.D.  $\times$  0.2 µm film thickness) with helium as the carrier gas. Injector and detector were set at 275°C. Samples (1 µL) were injected with a split ratio of 1:50 with a temperature program: 80°C for 4 min, then 15°C min<sup>-1</sup> to 270°C and finally 30°C min<sup>-1</sup> to 310°C (2 min). The peaks in the chromatograms corresponding to sugars and lactic acid were identified by their retention times. Quantifications were calculated using the peak areas and the calibration standard curve for each compound.

## Quantification of the 2-Substituted (1,3)-β-D-Glucan Produced by *P. parvulus*

A competition (ELISA) method for the specific detection of the EPS synthesized by *P. parvulus* 2.6, based on *Streptococcus pneumoniae* serotype 37 antibodies, was performed as previously described (Werning et al., 2014). Briefly, the ELISA assay was carried out in 96-Well Nunc-Immuno MicroWell MaxiSorp plates (Thermo Fisher Scientific), and the EPS of *P. parvulus* 2.6, purified as previously described (Notararigo et al., 2013), was immobilized in each well (62.5 ng per well).



Culture supernatants [diluted with phosphate-buffered saline (PBS) pH 7.2 when necessary] were used as competitor for binding to the primary antibody (dilution 1:800 of anti-serotype 37, Statens Serum Institut, Copenhagen, Denmark). Then, primary antibody was conjugated with a secondary antibody, polyclonal Anti-Rabbit IgG alkaline phosphatase (Sigma-Aldrich, Saint Louis, MO, United States) diluted 1:25,000, and finally was revealed with *p*-nitrophenylphosphate in diethanolamine buffer (Sigma-Aldrich). Reaction signals

were detected with a microtiter plate reader model 680 (Bio-Rad, Hercules, CA, United States), measuring the OD at 415 nm. Quantification was performed using a standard curve generated by the competition for the primary antibody of serial dilutions of the purified *P. parvulus* 2.6 EPS dissolved in PBS.

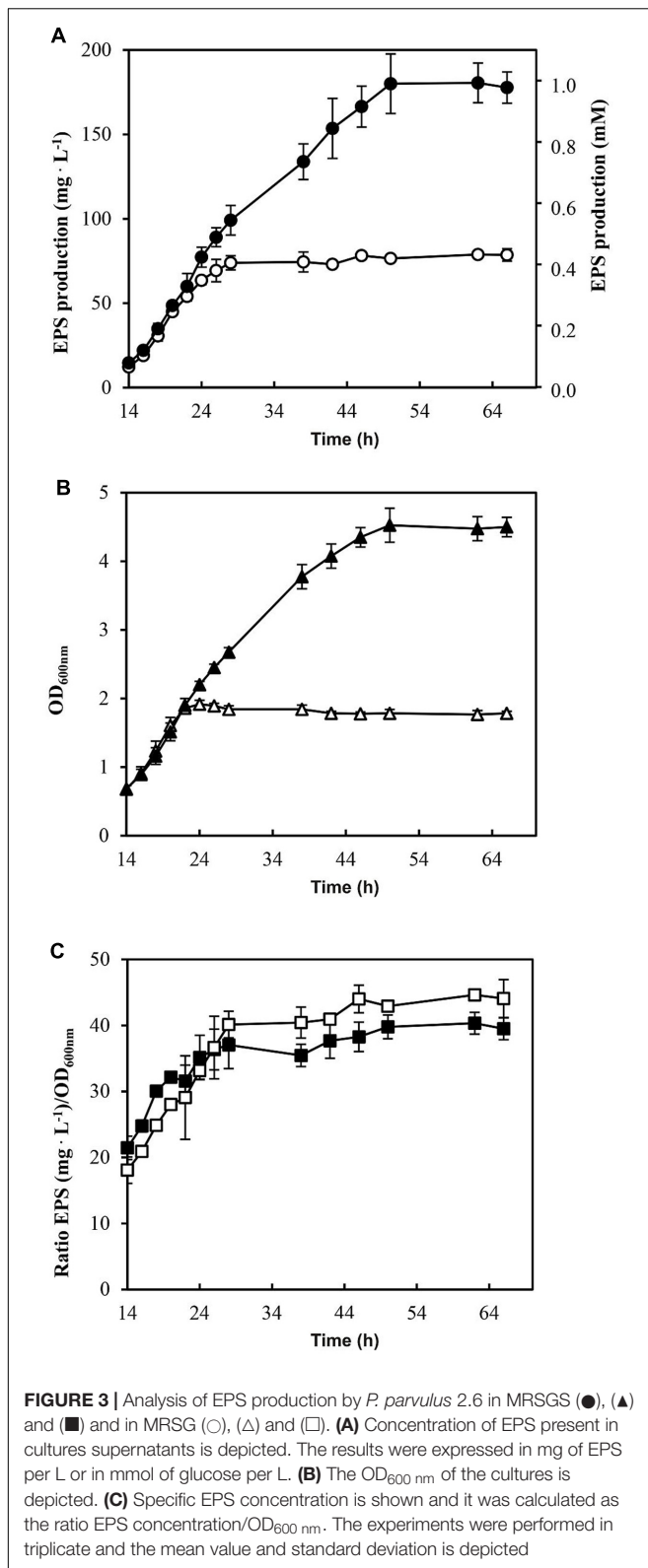
### Detection of mCherry Fluorescence in LAB Carrying pPCR16, pPCR17, pPCR18, or pPCR19

To detect the expression levels of the mCherry fluorescent protein, *L. plantarum* Lp90 strains carrying the pPCR derivatives were diluted 1:100 and grown in MRS supplemented with 1% glucose in static mode at 37°C, until mid-exponential phase. Then the cultures were centrifuged at 9,000 × *g* for 10 min at room temperature, and the cells were washed with one volume of PBS pH 7.2 prewarmed at 37°C. Then, the bacteria were resuspended in the same volume of MRS broth supplemented with 1% sorbitol or 1% glucose prewarmed at 37°C. Cultures were incubated at 37°C with agitation of 180 rpm, and samples were taken each hour. Two hundred microliter of all chilled samples were centrifuged at 9,000 × *g* for 10 min at 4 °C and cells were washed once with chilled PBS buffer pH 7.2. Samples were resuspended in 200 μL of PBS buffer pH 7.2 and used to measure the fluorescence levels of mCherry protein in a 96-Well Nunc U96 MicroWell plate (Thermo Fisher Scientific) in a Varioskan Flash equipment (Thermo Fisher Scientific), using 587 and 610 nm wavelengths for excitation and detection of emission, respectively. In addition, appropriate dilutions were prepared to estimate culture biomass by measuring the OD<sub>600 nm</sub>. Three independent trials were performed and the same fresh suspensions, without fixing, were used for phase contrast and fluorescent microscopy analysis with a Leica DM1000 model microscope (Leica Microsystems, Mannheim, Germany) with a light source EL6000 and a filter system TX2 ET for detection of red fluorescence. The microscope was connected to a DFC3000G camera (Leica Microsystems) with a CCD sensor. Image analysis was performed using Leica Application Suite X Software (Leica Microsystems).

To detect the expression of the mCherry fluorescent protein, *L. casei* BL23 strains carrying the pPCR derivatives were grown and processed in the same manner as the *L. plantarum* cultures, except that preinoculum cultures were diluted in MRS supplemented with 1% glucose or 1% sorbitol to an OD<sub>600 nm</sub> = 0.1 and then were incubated at 37°C with agitation of 180 rpm for 16 h, until they reached early stationary phase. Then, 1 mL of each culture was centrifuged and washed with PBS as above. Samples were concentrated five-fold and used to measure the fluorescence levels and to take fluorescence images as described above.

### Bioinformatic Analysis

The DNA sequence of plasmid pPP1 was analyzed with the programs included in the DNASTAR Lasergene 12 (DNASTAR Inc. Madison, WI, United States). Homologies of pPP1 DNA



sequences and of its inferred translated products with the NCBI data bases of the National Center for Biotechnology Information (NCBI) were analyzed with the Basic Local Alignment Search

Tool (BLAST)<sup>1</sup>. Multiple sequence alignment of genes and proteins were performed with Clustalx 2.1<sup>2</sup> programs.

Transmembrane helices in GutM were predicted using TMHMM 2.0<sup>3</sup> (TMPred<sup>4</sup>) programs. Prediction of secondary structures in the gut mRNA was accomplished with the mfold 2.3 program<sup>5</sup>.

## RESULTS

### Analysis of *P. parvulus* Sorbitol Metabolism

Sorbitol could be a substrate for the synthesis of *P. parvulus* 2.6 EPS and analysis of the DNA sequence of the draft genome of this bacterium (Pérez-Ramos et al., 2016) with the BLAST program revealed a putative *gut* operon, that could be involved in transport and catabolism of this compound. Therefore, growth of *P. parvulus* 2.6 and its isogenic EPS-non-producing (non-ropy) 2.6NR strain in MRS (without glucose) and MRSS (medium containing sorbitol) was tested. The 2.6NR strain showed the same poor growth in both media (Supplementary Figure S1). However, the presence of sorbitol in the medium significantly improved the growth of the 2.6 strain (Supplementary Figure S1), reaching a final OD<sub>600 nm</sub> of 3.0 in MRSS versus 0.45 in MRS, indicating that this bacterium was able to utilize sorbitol. Nevertheless, the growth of *P. parvulus* 2.6 in MRSS was very slow and took more than 12 days to reach the final optical density (Supplementary Figure S1). Therefore, in order to improve the growth rate of 2.6 strain, the influence of modifying various parameters in bacterial growth in MRSS was investigated. The best inferred conditions were the usage of a MRSGS containing as carbon sources 10 mM glucose plus 30 mM sorbitol, pH = 5.2, and growth with aeration at 30°C. Thus, these conditions were used to investigate a potential interplay between sorbitol utilization and EPS production by *P. parvulus* 2.6.

A comparative study of the metabolic fluxes of *P. parvulus* strains by analysis of culture supernatants during growth in MRSG or MRSGS corroborated that 2.6, but not 2.6NR, was able to ferment sorbitol (Figure 2). Co-metabolism of sorbitol and glucose by the 2.6 strain resulted in an increase of 2.5-fold in the final biomass estimated by the OD<sub>600 nm</sub> of the cultures. Values of  $4.48 \pm 0.18$  in MRGS (Figure 2A) compared to  $1.77 \pm 0.06$  reached in MRSG (Figure 2B), the latter being similar to  $1.36 \pm 0.05$  observed for the 2.6NR strain in MRGS (Figure 2C). In addition, a prolonged exponential growth phase of the 2.6 strain was observed in the MRSGS medium (50 h versus 20 h, Figures 2A,B). In the 2.6NR culture supernatants, the initial sorbitol levels (30 mM) remained constant during the entire time period of the assays, revealing that this bacterium was unable to transport sorbitol to the cytosol (Figure 2C). Moreover, the analysis of the carbon source consumption by the 2.6 strain

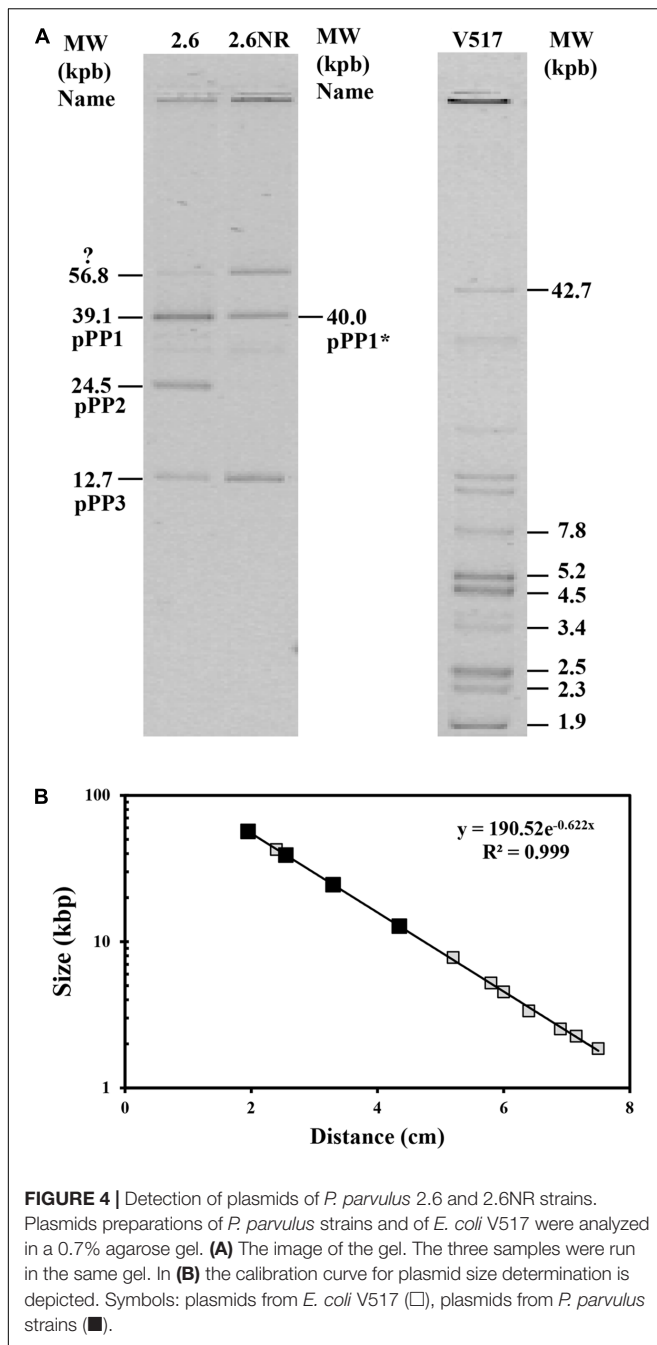
<sup>1</sup><https://blast.ncbi.nlm.nih.gov/Blast.cgi>

<sup>2</sup><http://www.ebi.ac.uk/Tools/msa/clustalw2/>

<sup>3</sup><http://www.cbs.dtu.dk/services/TMHMM-2.0/>

<sup>4</sup>[http://embnet.vital-it.ch/software/TMPRED\\_form.html](http://embnet.vital-it.ch/software/TMPRED_form.html)

<sup>5</sup><http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>



**FIGURE 4 |** Detection of plasmids of *P. parvulus* 2.6 and 2.6NR strains. Plasmids preparations of *P. parvulus* strains and of *E. coli* V517 were analyzed in a 0.7% agarose gel. (A) The image of the gel. The three samples were run in the same gel. In (B) the calibration curve for plasmid size determination is depicted. Symbols: plasmids from *E. coli* V517 (□), plasmids from *P. parvulus* strains (■).

showed that glucose started to be transported to the cytosol after 2 h of growth, and upon 26 h of incubation the monosaccharide was undetectable in the culture supernatants (Figures 2A,B). Furthermore, only after 20 h of incubation did the 2.6 strain start to internalize the sorbitol and presumably to metabolize it, because the bacterium did not enter into the stationary phase until the sorbitol was consumed (Figure 2A). The metabolic activity of the two strains was monitored by detecting the lactic acid production, since it is the main metabolic end-product because *pediococci* are homofermentative bacteria. The results showed that the 2.6 strain grown in MRSG (Figure 2B) and the

2.6NR strain grown in MRSGS (Figure 2C) released to the culture media similar amounts of lactic acid, the maximum levels being  $18.45 \pm 0.45$  mM and  $20.19 \pm 0.42$  mM, respectively. By contrast, the 2.6 strain grown in the presence of both carbon sources showed a higher lactic acid production, up to  $76.03 \pm 0.43$  mM (Figure 2A). Correlating with these results, the final pH of the 2.6 cultures in MRSG and of the 2.6NR cultures in MRSGS was similar ( $4.81 \pm 0.02$  versus  $4.78 \pm 0.02$ ), and higher than that of the 2.6 cultures in MRSGS ( $4.23 \pm 0.02$ ).

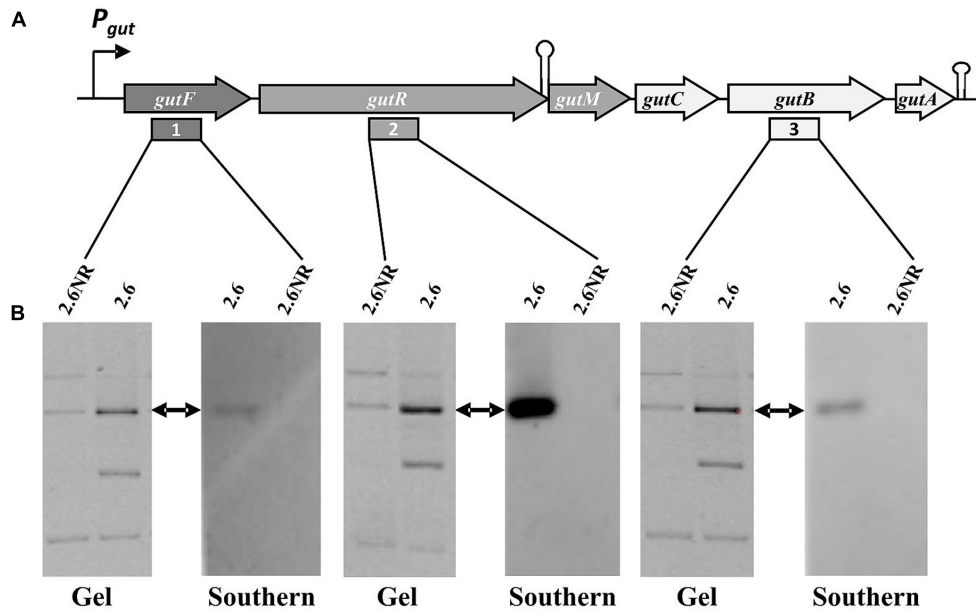
Furthermore, the EPS production by *P. parvulus* 2.6 in the presence or absence of sorbitol was investigated. Significant EPS levels were detected after 14 h of growth in MRSG and MRSGS media. Therefore, the data depicted in Figure 3 correspond to those obtained within the 14–62 h incubation period. The results revealed that the bacterium produced EPS during the growth in MRSGS and synthesized higher levels of the polymer in this medium than in MRSG (Figure 3A). Thus, after 62 h of growth in MRSG, the 2.6 strain produced  $78.6 \pm 3.7$  mg L<sup>-1</sup> of EPS, while in MRSGS synthesized  $180.5 \pm 11.8$  mg L<sup>-1</sup>. Additionally, in order to evaluate the specific efficiency of the EPS production depending on the carbon source used, the ratio between EPS concentration and the biomass estimated from the OD<sub>600 nm</sub> (Figure 3B) was calculated (Figure 3C). The results showed that irrespectively of the carbon source, the bacteria had almost identical efficiency of EPS production, which increased during the exponential and stationary phases of growth (Figure 3C).

## Determination of Genomic Location of the *gut* Operon

*P. parvulus* 2.6 probably carries three natural plasmids, which were previously named pPP1, pPP2 and pPP3 (Werning et al., 2006), and we have identified only three plasmid replication machineries in the *P. parvulus* draft genome (Pérez-Ramos et al., 2016). In addition, the *P. parvulus* 2.6 EPS is synthesized by the GTF glycosyltransferase encoded by the *gtf* gene, which is located in the pPP2 plasmid (Werning et al., 2006). Thus, the 2.6NR strain was generated from 2.6 by pPP2 plasmid curing after treatment with the DNA intercalating agent ethidium bromide and the gyrase inhibitor novobiocin (Fernández et al., 1995).

Consequently, given that 2.6NR does not utilize sorbitol, it was feasible that the *gut* operon was encoded by pPP2 and this hypothesis was investigated. First, total plasmidic DNA preparations of the two *Pediococcus* strains were purified by fractionation in a CsCl gradient to eliminate non-supercoiled (open circles and linear) forms of the plasmids. Then, the purified plasmidic DNA preparations were analyzed in an agarose gel (Figure 4). Four and three bands were detected, respectively, in preparations of the 2.6 and 2.6NR strains. The sizes of the bands were inferred from their migration using a calibration curve (Figure 4B) generated with the plasmids of the *E. coli* V517 strain and are shown in Figure 4A. Two of the bands apparently were shared by 2.6 and 2.6NR, and were initially ascribed to the monomeric forms of pPP1 (39.1 kpb in 2.6 and 40.0 kpb in 2.6NR) and pPP3 (12.7 kpb). As expected, pPP2 (24.5 kpb) was not detected in 2.6NR DNA preparations. Moreover, we could not ascribe to any plasmid the band with less mobility and a





**FIGURE 5 |** Detection of the *P. parvulus* 2.6 *gut* operon by Southern blot hybridization. Plasmids preparations of the *P. parvulus* strains were fractionated by electrophoresis in agarose gel, transferred to membranes and hybridized for detection of *gutF*, *gutR* and *gutB* with the probes 1, 2, and 3, respectively. **(A)** Physical map of the *gut* operon, including the putative promoter of the *gut* operon ( $P_{gut}$ ), the operon genes (*gutF*, *gutR*, *gutM*, *gutC*, *gutB* and *gutA*), a secondary structure including the 3'-end of *gutR* and the 5'-end of *gutM* (coordinates 3483-3512 in GenBank accession No MF766019) as well as a transcriptional terminator (coordinates 6103-6027 in GenBank accession No MF766019) downstream of *gutA*. In addition, location of probes 1, 2, and 3 is indicated. **(B)** Images of the corresponding gels and hybridized membranes. The double headed arrows indicate the position of the hybridized bands in the corresponding gel.

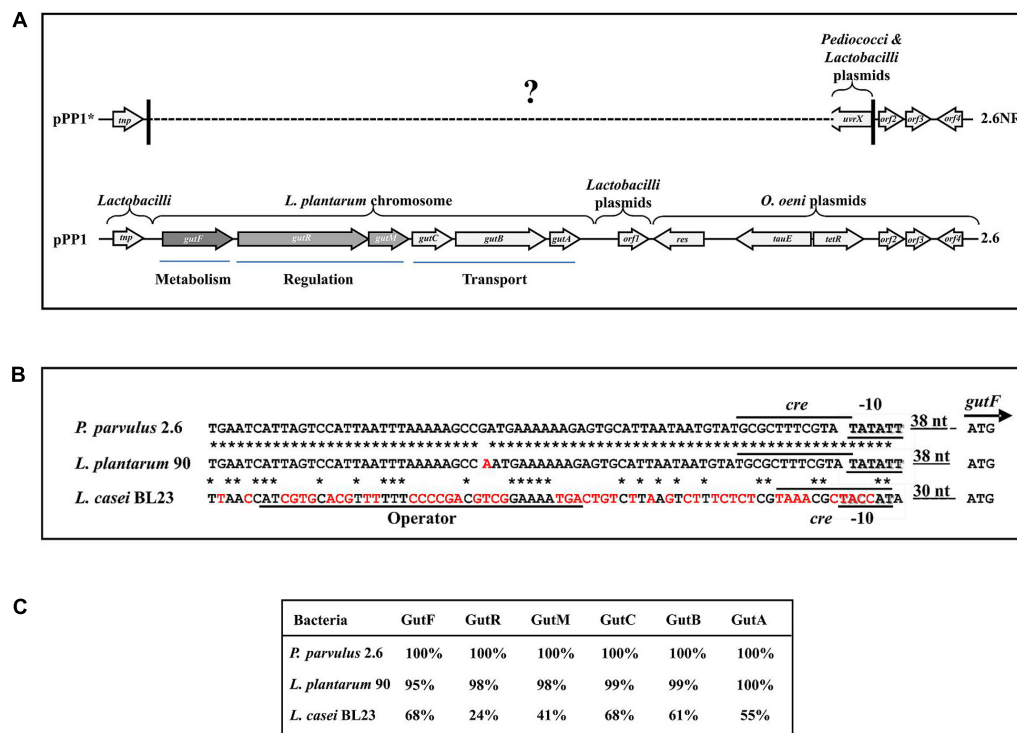
theoretical molecular weight of 56.8 kbp that was present in DNA preparations of both strains. Quantification of the bands from agarose gels (Figures 5, 6) revealed different proportions of the plasmidic forms in 2.6 (0.3:5.9:2.5:1.0) and 2.6NR (0.8:1.0:0.0:1.0) samples.

The *gut* operon of *P. parvulus* 2.6 (Figures 5A, 6A) is composed of six genes, of which *gutF* encodes a sorbitol-6-phosphate dehydrogenase; *gutRM* encodes two putative regulators; and *gutCBA* encodes the proteins EIIC, EIIBC and EIIA which are components of a phosphoenolpyruvate-dependent sorbitol phosphotransferase system (PTS<sup>gut</sup>). Thus, to detect the location of the *gut* operon, Southern blot hybridization of total plasmidic DNA preparations was performed using as a probe internal regions of *gutF*, *gutR* or *gutB*. One hybridization signal was observed with the three probes at the position of the 39.1 kb pPP1 plasmid in the 2.6 DNA sample (Figure 5B). Surprisingly, this plasmid was apparently present in both *P. parvulus* strains, but in the 2.6NR DNA sample no signal was observed. Nevertheless, the results demonstrated that the *gut* operon was not located in the pPP2 plasmid, but rather was carried by the pPP1 plasmid of the 2.6 strain and not of the newly designated pPP1\* plasmid of 2.6NR strain.

### Analysis of Plasmids pPP1 of *P. parvulus* 2.6 and pPP1\* of *P. parvulus* 2.6NR

The results obtained by Southern blot analysis prompted us to obtain further information of pPP1 and pPP1\* plasmids. Thus, the total plasmidic DNA preparation of the 2.6 strain

was used as a substrate to confirm the sequence of the *gut* operon and to determine the unknown nucleotide sequence of the flanking regions (undetected in the draft genome of the bacterium) by the dideoxynucleotide method and with the walking strategy. The sequence of a DNA segment of 11,746 bp (Figure 6A and GenBank accession No MF766019) was obtained and its analysis revealed the existence of nine open reading frames (ORF), in addition to the 6 genes (*gutFRMCBA*) of the *gut* operon (Figure 6A and Supplementary Table S1). One open reading frame was detected upstream of the *gut* operon and was designated *tnp*, since its product has 100% identity with a multispecies transposase (Genbank accession No WP\_003606336.1) widely distributed in the Lactobacillaceae family. Downstream of the *gut* operon were detected four ORF named *orf1*, *orf2*, *orf3* and *orf4*, which could encode hypothetical proteins conserved in other LAB. In addition, the product of the named *res* gene belongs to the Ser-recombinase superfamily (cl02788) and specifically to the PinE conserved protein domain family (COG1961), showing more than 90% amino acid identity with proteins from *Oenococcus*, *Lactobacilli* and *Pediococci* annotated as Pin-related site-specific recombinases/DNA invertases. Also, two divergent genes named *tauE* and *tetR* seem to encode a TauE sulfite exporter which belongs to the TauE conserved domain family (pfam01925) and a transcriptional regulator belonging to the TetR family (domain architecture ID 11442015), and both proteins have more than 95% amino acid identity with their homologues in *Oenococcus oeni* and *Lactobacilli*.



**FIGURE 6 |** The sorbitol operon and flanking DNA regions of *P. parvulus* 2.6. **(A)** Sequenced regions of pPP1 and pPP1\* plasmids are shown. Genes and ORF are depicted and homologies at the nucleotide level are indicated (see details in the text). ? Indicates unknown sequence. Homology of regulatory genetic regions **(B)** and proteins **(C)** involved in sorbitol utilization between *P. parvulus* 2.6, *L. plantarum* 90 and *L. casei* BL23 is shown. **(B)** DNA sequences of: the putative  $P_{gut}$  promoters (–10 boxes) and the GutR operator of *L. casei* are underlined; the putative *cre* sequences for catabolic repression are overlined; also the start codon (ATG) of *gutF* gene and the nucleotides between this and the –10 boxes are depicted. Red letters indicate mismatches with the *P. parvulus* 2.6 DNA sequence. \*Means identity of nucleotides. **(C)** Identity of the amino acid sequence of the Gut proteins from *L. plantarum* 90 and *L. casei* BL23 with those of *P. parvulus* 2.6 is depicted.

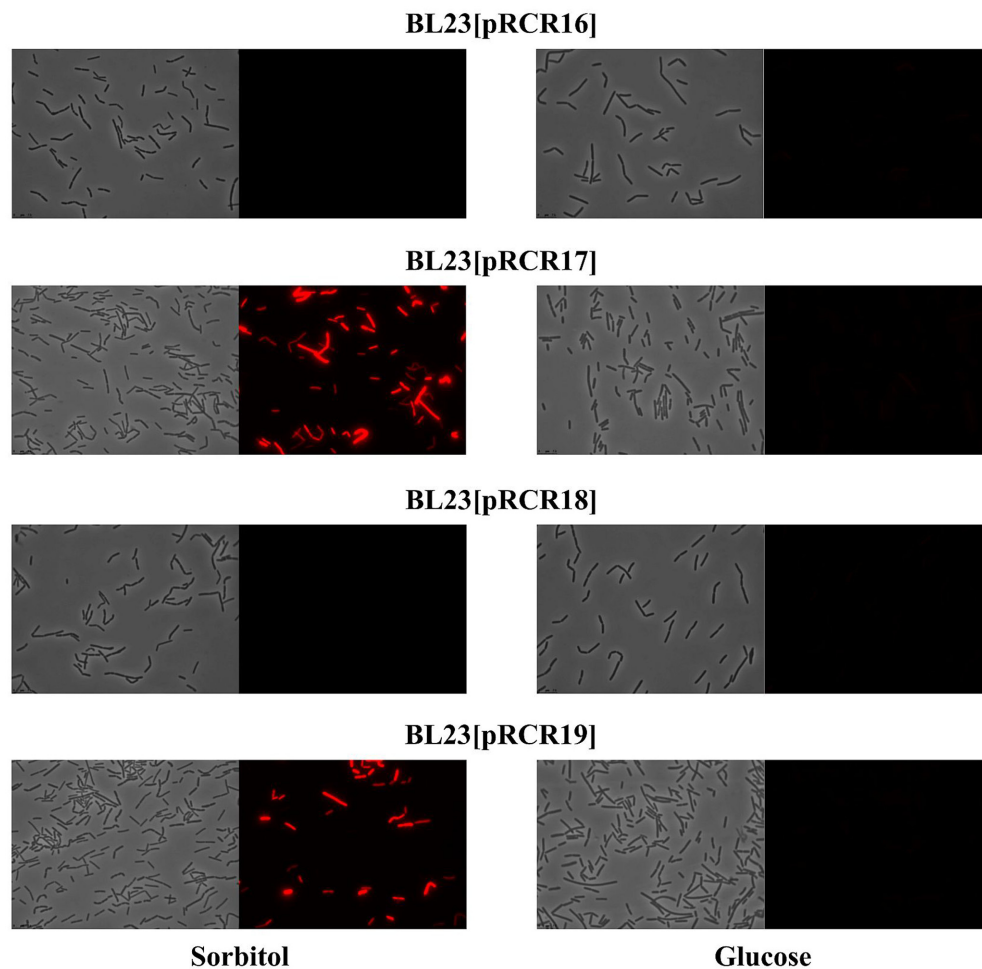
Based on the DNA sequence of the *gut* operon flanking regions in pPP1, and on the lack of the *gut* operon in 2.6NR, primers were designed and used to try to detect if there exists any identity between pPP1 and pPP1\* by DNA sequencing. Two of these, pPP1\*F and pPP1\*R, located respectively upstream and downstream of the *gut* operon, provided the desired information (**Figure 6A**). A good chromatogram of the DNA sequencing of pPP1\* using the 2.6NR plasmidic preparation and pPP1\*F primer with 100% identity with pPP1 was obtained until nucleotide 156 in the chromatogram (548 nt in Genbank accession No WP\_003606336.1), then at least two overlapping sequences were observed (Supplementary Figure S2A), and it was not possible from this point to deduce a further correct DNA sequence. This was not the case when DNA from the 2.6 strain was used as substrate, since a good chromatogram of the pPP1 DNA sequencing was obtained (Supplementary Figure S2B). However, the usage of pPP1\*R allowed not only to determine that the homology between pPP1 and pPP1\* starts again at nucleotide 10,021 (in Genbank accession No WP\_003606336.1), but also that upstream of this position in pPP1\* there exists a region including a *uvrX* putative gene identical to those of other pediococci (i.e., in pPC892-2 plasmid, Genbank accession No CP021472.1) and *Lactobacilli* (i.e., in pH10 plasmid, Genbank accession No

CP002430.1) plasmids, which do not carry *orf2*, *orf3* and *orf4*.

With regard to the *gut* operon of *P. parvulus* 2.6, the identity of the region including the genes and the upstream regulatory regions with the homologues of *L. plantarum* strains was 99% (**Figures 6A,B**) and nucleotides from 694-to 60020 in GenBank accession No MF766019), consequently the amino acid sequence of the Gut proteins of *P. parvulus* showed an identity ranging from 95 to 100%, with those of *L. plantarum* 90 (**Figure 6C**). No significant homology at the DNA sequence level was detected between the characterized operons of *L. casei* and those of *P. parvulus* (**Figure 6B** and results not shown). However, presumably due to convergent evolution, homology ranging from 68 to 24% amino acid identity was detected between the Gut proteins of *P. parvulus* 2.6 and of *L. casei* BL23 (**Figure 6C**).

## Analysis of the Gut Operon Regulation

GutR and GutM of *P. parvulus* could be involved in regulation of the *gut* operon expression and upstream of the start codon of *P. parvulus* 2.6 *gutF* gene, a TATAAT sequence was detected that only deviates one nucleotide from the consensus –10 promoter region (**Figure 6B**). Thus, to gain insight into this potential regulation, complementation studies in heterologous LAB hosts able to utilize sorbitol were carried out. First, we cloned



**FIGURE 7 |** Detection of fluorescence in *L. casei* BL23 strains at stationary phase. Cultures of the indicated strains in MRSS (sorbitol) or MRSG (glucose) were analyzed by phase contrast (left panels) or fluorescence (right panels) microscopy.

independently the putative promoter sequence (designated  $P_{gut}$ ) and its upstream region (**Figure 6B**), as well as the transcriptional fusions  $P_{gut-gutR}$ ,  $P_{gut-gutM}$  and  $P_{gut-gutRM}$  into the pRCR promoter probe vector (Mohedano et al., 2015) upstream of the *mrfp*, generating the pRCR16, pRCR17, pRCR18 and pRCR19 plasmids, respectively (**Figure 1**). Thus, functionality of the promoter and influence of GutR and GutM could be detected by measuring the levels of fluorescence of the mCherry encoded by the *mrfp* gene. As hosts to perform the studies, we chose: (i) the plasmid free *L. casei* BL23, because its sorbitol utilization and the regulation of its *gut* operon is known (Yebra and Pérez-Martínez, 2002; Nissen et al., 2005; Alcantara et al., 2008) and, (ii) *L. plantarum* 90, because we have previously detected in this bacterium efficient functional expression of mCherry from a pRCR derivative, without problems of plasmid incompatibility and that the copy number of the plasmid was  $62 \pm 2$  molecules per bacterial genome (Russo et al., 2015).

The well characterized transcriptional activator GutR of *L. casei* BL23 controls expression of the *gut* operon of this bacteria and its operator site upstream of the  $P_{gut}$  has been identified as

well as a catabolite repression element (*cre*) overlapping the  $-10$  region of the promoter (Alcantara et al., 2008) (**Figure 6B**). The *P. parvulus* 2.6 GutR has only a low homology of amino acids (24%) with its homologue of *L. casei*, but like its counterpart belongs to the BglG transcriptional antiterminators family, possesses the PRD domain and the DNA helix turn helix binding domain. Therefore, both proteins could have a similar role. Alignment of the *L. casei* and *P. parvulus*  $-10$  regions revealed that the upstream regulatory regions of BL23 strain has no clear homologs in the 2.6 strain (**Figure 6B**). Consequently, cross talk between transcriptional signals of *P. parvulus* and *L. casei* regulators should not take place, and influence of the pediococcal GutR and GutM in expression of  $P_{gut}$  from the 2.6 strain could be investigated in the BL23 strain without interferences. Thus, the pRCR derivatives were transferred independently to the BL23 strain and the recombinant bacteria were grown in MRS supplemented with either 1% glucose or 1% sorbitol until stationary phase prior to analysis. Examination of the cultures by fluorescent and phase contrast optical microscopy revealed that only bacteria carrying pRCR17 and pRCR19 and grown in

**TABLE 3 |** Heterologous expression of components of the *P. parvulus* 2.6 *gut* operon in *L. casei* BL23 carrying pPCR derivatives plasmids grown in either MRSS or MRSG.

<i>L. casei</i> strains	DNA insert in pPCR derivatives	Specific fluorescence <sup>a</sup>	
		MRSS	MRSG
BL23[pPCR16]	<i>P<sub>gut</sub></i>	0.14 ± 0.07	0.14 ± 0.01
BL23[pPCR17]	<i>P<sub>gut</sub>-gutR</i>	5.69 ± 0.44	0.07 ± 0.09
BL23[pPCR18]	<i>P<sub>gut</sub>-gutM</i>	0.19 ± 0.06	0.11 ± 0.02
BL23[pPCR19]	<i>P<sub>gut</sub>-gutRM</i>	3.58 ± 0.06	0.05 ± 0.02

<sup>a</sup>The specific fluorescence is depicted and it was calculated as the ratio of the detected fluorescence (5×) and the bacterial biomass estimated from the OD<sub>600 nm</sub> of the culture.

presence of sorbitol have fluorescence (Figure 7). In addition, fluorescence as well as the optical density of the cultures was measured and the specific fluorescence, referred to the biomass, was calculated. The fluorescence quantification confirmed that the *P<sub>gut</sub>-gutRmrfp*, and *P<sub>gut</sub>-gutRMmrfp* transcriptional fusions are activated upon growth in the presence of sorbitol (Table 3). Thus, these results revealed that expression from the *P<sub>gut</sub>* required the activation by GutR and the presence of sorbitol in the growth medium. Moreover, they indicated that activation by GutR decreased, when GutM was present (5.69 ± 0.44 versus 3.58 ± 0.06).

Concerning the *L. plantarum* 90 host, its GutR has 98% homology to that of *P. parvulus* 2.6 (Figure 6C) and the DNA sequence of the region located upstream of the two *P<sub>gut</sub>* promoters only differs in one nucleotide (Figure 6B). Consequently, both operons must have the same regulatory gene system, which implies that both systems could recognize each other. Thus, a trans-complementation process was expected between the regulatory proteins of Lp90 and the promoter region of 2.6. Therefore, the pPCR derivatives were transferred independently to the 90 strain and, since a cross talk is more complex situation, its comprehension required a more detailed analysis. For this reason, the recombinant bacteria, after growth in MRS supplemented with 1% glucose, were transferred to MRS fresh medium supplemented with either 1% sorbitol or 1% glucose and a time course assay of fluorescence and growth of the cultures was performed. The results revealed that all recombinant strains became fluorescent, when grown in the presence of sorbitol and, with the time of incubation the fluorescence increased (Figure 8 and Table 4). In addition, analysis of the bacterial growth showed that all cultures in MRSG have very similar exponential growth rates (ranging from 0.889 ± 0.059 to 0.803 ± 0.049) and all entered slowly into stationary phase after 2 h of incubation (Figure 8F and Table 4). Initial transfer of the cultures to MRSS resulted in a similar decrease (around 50%) of the growth rate (values from 0.416 ± 0.045 to 0.495 ± 0.011) during the first 2 h of induction. Then, probably after consumption of the residual intracellular glucose or due to the induction process, bacteria decreased their growth rate to levels ranging from 0.259 ± 0.020 to 0.251 ± 0.034, besides the 90[pPCR18] (GutM overexpressor), that after stalling its growth from 2 h to 3 h incubation time

decreased its growth rate to 0.239 ± 0.048, indicating that overexpression of GutM in absence of high levels of GutR has a negative impact for the cells. Furthermore, analysis of the specific levels of fluorescence of the cultures referred to their biomass (Table 4) showed different levels for the different fusions (*P<sub>gut</sub>-mrfp* < *P<sub>gut</sub>-gutRMmrfp* < *P<sub>gut</sub>-gutRmrfp* < *P<sub>gut</sub>-gutMmrfp*), showing that overexpression of GutM provokes the highest induction of expression from *P<sub>gut</sub>*. In addition, the highest levels were observed after 4 h of induction for cells carrying either pPCR17 (22.38 ± 2.02) or pPCR19 (19.26 ± 2.10) versus the end of the incubation (6 h) for cells carrying pPCR16 (14.81 ± 0.66) and pPCR18 (26.83 ± 1.83).

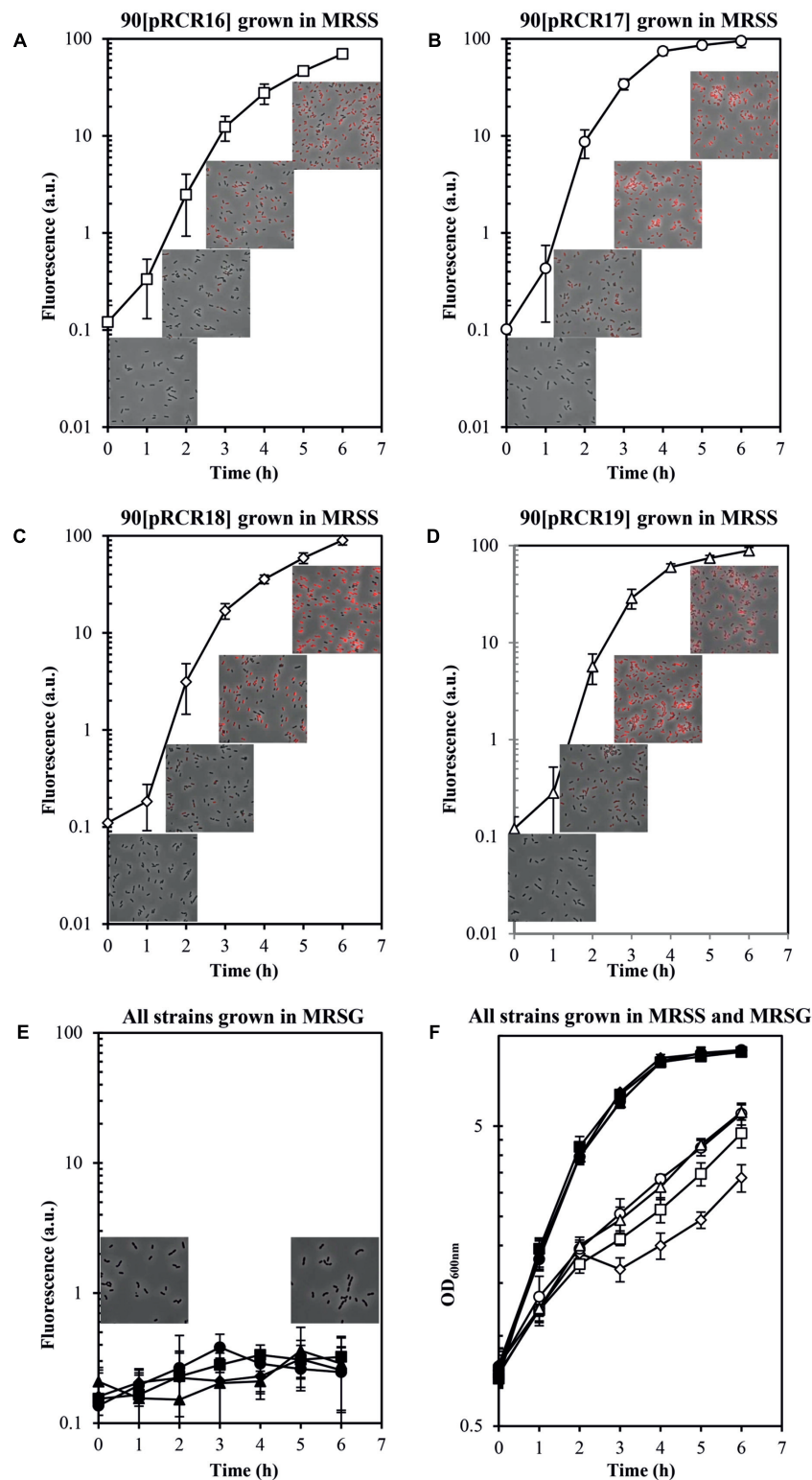
Thus, the results revealed a trans-complementation of the *L. plantarum* regulatory proteins on expression driven from the *P. parvulus* *P<sub>gut</sub>* promoter. Moreover, the results confirmed the role of inducer of GutR as well as requirement of sorbitol for expression from *P<sub>gut</sub>* and support that co-expression of GutR and GutM decrease the activation mediated by GutR.

## DISCUSSION

The overall metabolic results obtained here support that *P. parvulus* is able to synthesize EPS in MRS medium using either glucose or sorbitol as carbon sources. We have previously demonstrated that the 2-substituted (1,3)-β-D-glucan of *P. parvulus* 2.6 is synthesized by the GTF glycosyltransferase utilizing UDP-glucose as substrate (Werning et al., 2014). In addition, Velasco et al. (2007) determined that the 2.6 strain transport the glucose by a PMF-permease and possesses the α-phosphoglucosyltransferase and the UDP-glucose pyrophosphorylase activities responsible for the conversion of glucose-6-P to glucose-1-P and further conversion of this compound to UDP-glucose. Thus, Velasco et al. (2007) showed how the 2.6 strain uses the glucose, not only for the central metabolism, via the glycolytic pathway, but also for the secondary metabolism involving a biosynthetic pathway for its EPS synthesis. In addition, the detection of the genetic determinants of sorbitol utilization by the 2.6 strain obtained in this work supports that the bacterium transports sorbitol by a PTS<sup>Gut</sup> system and converts sorbitol-6-P into fructose-6-P by the action of sorbitol-6-P dehydrogenase. Fructose-6-P can be converted to glucose-6-P by a reaction catalyzed by phosphoglucose isomerase, enzymatic activity that was also previously detected in the 2.6 strain (Velasco et al., 2007). Therefore, the 2.6 strain possesses the transport and enzymatic machineries for synthesis of the EPS from sorbitol. In addition, we have detected that aeration of the cultures during the growth improves sorbitol consumption (results not shown). Accordingly, the conversion of sorbitol-6-P into fructose-6-P requires NAD<sup>+</sup> as an oxidative co-factor to produce NADH (Zarour et al., 2017). Analysis of the draft genome of the 2.6 strain showed the existence of a putative NADH oxidase coding gene. If this enzyme exists, it could unbalance the NAD<sup>+</sup>/NADH equilibrium toward the oxidized form NAD<sup>+</sup>.

The *P. parvulus* 2.6 2-substituted (1,3)-β-D-glucan is composed of molecules of glucose and consequently the





**FIGURE 8 |** Trans complementation of the sorbitol utilization regulatory machinery in *L. plantarum* 90. Bacteria carrying **(A,E,F)** pRCR16 (□, ■), **(B,E,F)** pRCR17 (○, ●), **(C,E,F)** pRCR18 (◇, ◆) or **(D,E,F)** pRCR19 (△, ▲) were grown in MRSG and at time 0 were transferred to fresh MRSS (open symbols) or to MRSG (closed symbols).  $OD_{600nm}$  and fluorescence of the cultures were monitored every hour. Overlays of images of the cultures taken at time 0 and after 2, 4, and 6 h incubation by phase contrast and fluorescent microscopy are depicted. The experiments were performed in triplicate and the mean value and standard deviation is depicted.

TABLE 4 | Heterologous expression of components of *P. parvulus* 2.6 *gut* operon in *L. plantarum* 90 carrying pCRC derivatives plasmids.

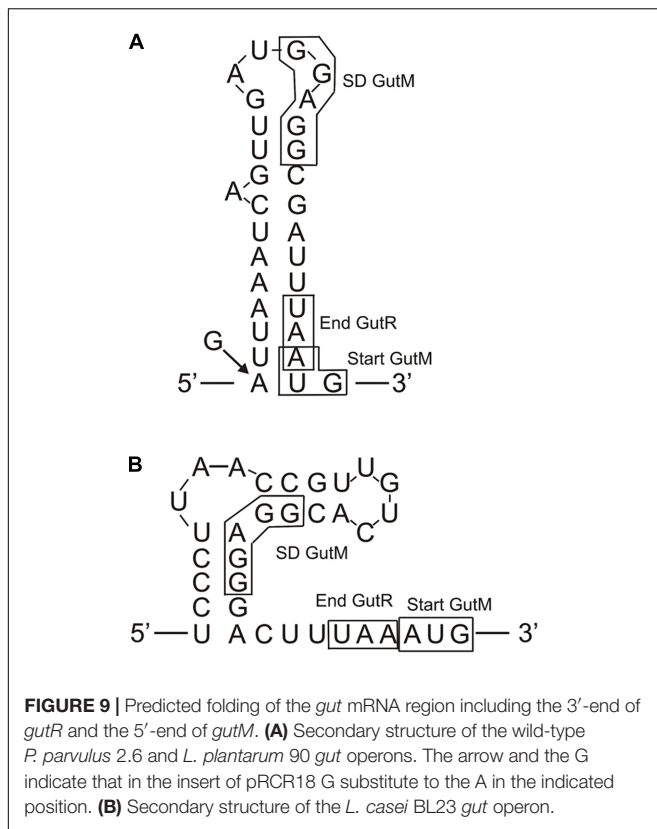
<i>L. plantarum</i> strains	DNA insert in pRCR derivatives	Medium	Initial growth rate <sup>a</sup> (μ)	Sorbitol induced growth rate <sup>b</sup> (μ)	Specific fluorescence <sup>c</sup>				
					2 h	3 h	4 h	5 h	6 h
90[pRCR16]	<i>P<sub>gut</sub></i>	MRSS	0.416 ± 0.045	0.251 ± 0.034	1.48 ± 1.03	5.86 ± 1.43	10.42 ± 1.56	13.44 ± 0.51	14.81 ± 0.66
90[pRCR17]	<i>P<sub>gut</sub>-gutR</i>	MRSS	0.455 ± 0.065	0.259 ± 0.020	4.42 ± 1.22	13.41 ± 0.90	22.38 ± 2.02	20.30 ± 1.62	17.28 ± 1.65
90[pRCR18]	<i>P<sub>gut</sub>-gutM</i>	MRSS	0.429 ± 0.025	0.239 ± 0.048	1.62 ± 0.69	10.26 ± 2.30	18.05 ± 2.85	24.26 ± 1.51	26.83 ± 1.83
90[pRCR19]	<i>P<sub>gut</sub>-gutRM</i>	MRSS	0.495 ± 0.011	0.256 ± 0.016	2.80 ± 0.77	11.74 ± 1.51	19.26 ± 2.10	17.20 ± 1.71	15.98 ± 1.21
90[pRCR16]	<i>P<sub>gut</sub></i>	MRSG	0.889 ± 0.059	<i>ND</i>	0.06 ± 0.03	0.04 ± 0.00	0.04 ± 0.00	0.04 ± 0.01	0.04 ± 0.01
90[pRCR17]	<i>P<sub>gut</sub>-gutR</i>	MRSG	0.825 ± 0.029	<i>ND</i>	0.07 ± 0.06	0.06 ± 0.02	0.04 ± 0.01	0.03 ± 0.00	0.03 ± 0.02
90[pRCR18]	<i>P<sub>gut</sub>-gutM</i>	MRSG	0.803 ± 0.049	<i>ND</i>	0.06 ± 0.03	0.03 ± 0.03	0.03 ± 0.01	0.04 ± 0.01	0.03 ± 0.01
90[pRCR19]	<i>P<sub>gut</sub>-gutRM</i>	MRSG	0.832 ± 0.034	<i>ND</i>	0.04 ± 0.02	0.03 ± 0.02	0.03 ± 0.01	0.04 ± 0.02	0.03 ± 0.02

<sup>a</sup>Growth rate of the cultures grown in MRSS was calculated from the OD<sub>600 nm</sub> obtained during the first 2 h of induction. <sup>b</sup>Growth rate of the cultures grown in MRSS was calculated from the OD<sub>600 nm</sub> obtained during the 2–6 h of induction, besides for 90[pRCR18], that due to the stalling of growth from 2 h to 3 h of incubation the growth rate was calculated from the data obtained from 3 h to 6 h of incubation. ND, the growth rate was not determined because the cultures have entered in the stationary phase of growth. <sup>c</sup>Specific fluorescence was calculated as the ratio of the detected fluorescence (5x) and the bacterial biomass estimated from the OD<sub>600 nm</sub> of the culture.

EPS concentration can be calculated as molarity of this monosaccharide (see secondary Y axis in **Figure 3A**). This calculation revealed that in both media this bacterium only used a small percentage of the substrate molecules (10 mM glucose plus 30 mM sorbitol in MRSGS or 10 mM glucose in MRSG) for synthesis of EPS (0.99 or 0.45 mM, respectively), whereas more than 90% was utilized in the glycolytic pathway to synthesize pyruvic acid (2 molecules per 1 molecule of substrate) and by action of the lactate dehydrogenase to finally generate lactic acid (1 molecule per 1 molecule of pyruvate, 79 mM or 18 mM). Moreover, the specific quantification method for 2-substituted (1,3)-β-D-glucan used here and the estimation of the specific concentration of EPS synthesized (**Figure 3C**) showed that, using as substrate either glucose or sorbitol, the bacterium synthesizes the same polymer and suggests that with the same efficiency. This was not the case when synthesis of this EPS utilizing fructose was tested, since levels were low compared with that obtained from glucose (Velasco et al., 2007). We have also observed a temporal delay of 2.6 to start to consume sorbitol in MRSGS (**Figure 3A**). This could be due to the existence of a catabolite repression of sorbitol utilization by glucose. Supporting this hypothesis, we have detected a potential *cre* operator (**Figure 6**) for the CcpA, which mediates with HPR this regulation in firmicutes (Deutscher, 2008).

In *P. parvulus* 2.6, the *gut* operon, as in other LAB, constitutes the genetic determinant for sorbitol transport and conversion into fructose-6-P. In addition, we have established here that it is located in a plasmid named pPP1 (**Figure 6**) which is unusual, since the almost identical operon of *L. plantarum* and that of *Lactobacillus pentosus* strain SLC13 (82% homologous, Genbank accession No CP022130.1) as well as the unrelated one from *L. casei* are located in the chromosome. As far as we know, only the location of an unrelated *gut* operon in the megaplasmid pMP118 from *L. salivarius* UCC118 has been previously described (Claesson et al., 2006). A search of the protein data banks revealed that *L. salivarius* 5713 and JCM1046 strains possess, respectively, the pHN3 and pMP1046A megaplasmids which carry *gut* operons homologous to that of pMP118 (Jiménez et al., 2010; Raftis et al., 2014). Flanking the operon two inverted repeat sequences (nucleotides 604–627 and 6612–6635 of Genbank accession No WP\_003606336.1) were identified, which are also present at the same relative location in *L. plantarum* strains and at various locations in lactobacilli chromosomes and plasmids (even more than one copy per genome). The upstream region is preceded by a *tnp* gene encoding a putative transposase, which could be responsible for mobilization of the *gut* operon from plasmid to chromosome or vice versa.

Lactic acid bacteria are prone to carry more than one compatible plasmid and this facilitates exchange of different regions with physiological significance, that later on can be transferred to other bacteria by plasmid conjugation or mobilization (Cui et al., 2015). Thus, downstream of the *gut* operon of pPP1 there are DNA regions almost identical to that present in plasmids of lactobacilli, which along with *P. parvulus* can be contaminants of alcoholic beverages. Furthermore, the



*Oenococcus oeni* pOENI-1 and pOENI-1v2 plasmids (Favier et al., 2012) and pPP1 carry a region containing among others the *res*, *tauE* and *tetR* genes. The putative TauE sulfite exporter is possibly involved in adaptation to stress conditions during alcoholic beverage production (Favier et al., 2012). Thus, the recombinase or invertase site specific *Res* could be responsible for a mobilization of an element composed of a truncated *res*, *tauE* and *tetR* to a stable location, since at the 3'-end region of *res* and downstream of *tetR* unit exist inverted complementary sequences 5'-TTTTAAAGC-3' and 5'-GCTTTAAAA-3' (nucleotides 7774–7778 and 10021–10029 of Genbank accession No WP\_003606336.1).

Another instance of plasmids rearrangement in *P. parvulus* is that which generated the profile and DNA sequence of 2.6NR strain plasmids (Figure 6). The initial isolate of 2.6NR strain generated in the Basque country University (BCU, Spain) and described in Fernández et al. (1995), kindly provided by Dr. Maria Teresa Dueñas (BCU) was studied in this work. Thus, the changes in plasmid cassettes were not produced in our laboratory, and presumably they took place upon treatment of 2.6 strain with ethidium bromide and novobiocin and was selected for the loss of the ropy phenotype. Thus, it is feasible that a formation of a co-integrate of pPP1 with other plasmid, may be pPP2, took place and convergent replication from two origins prompted to a deletion of one of the replicons, may be the pPP2, since its loss was envisaged, to generate pPP1\*.

Concerning the regulation of expression of the *gut* operon, the overall results showed that it is repressed in the absence

of sorbitol in the growth medium and that the *P. parvulus* GutR is an activator like the *L. casei* BL23 regulator (Alcantara et al., 2008). In this system, it has been proposed that GutM is involved in the activation, since a decreased expression of the *gut* operon was detected in a GutM deficient mutant (Alcantara et al., 2008). Furthermore, the complementation studies in *L. plantarum* 90 performed here showed a heterologous regulation of gene expression from the pediococcal  $P_{gut}$  promoter by the GutR from *Lactobacillus*, and a positive effect when only the pediococcal GutM was overexpressed (Figure 8 and Table 4). Thus, these results suggest that a protein-protein interaction between the *P. parvulus* GutM and the *L. plantarum* GutR could potentiate the activation of the  $P_{gut}$  promoter, since, *P. parvulus* 2.6 GutR and GutM have 98% identity with their homologues of *L. plantarum* 90. In addition, either in *L. casei* BL23 and in *L. plantarum* a decrease of expression from  $P_{gut}$  was observed when GutM was overexpressed in combination with GutR (Figure 8 and Tables 3, 4). This prompted us to analyze the genetic environment of *gutR* and *gutM*. An overlapping of the last nucleotide of the termination codon (TAA) of *gutR* and the first nucleotide of the start codon (ATG) of *gutM* was detected in *P. parvulus* 2.6 and *L. plantarum* 90 genomes. This indicated that post-transcriptional regulation of the *gut* operon could exist in this bacterium. For this reason, the secondary structure of the region surrounding the overlapping in the *gut* mRNA was folded with the Mfold program (Figure 9A). A secondary stem-loop structure was predicted with a  $\Delta G = -5.6 \text{ kcal mol}^{-1}$ , the ribosomal binding site (RBS) of *gutM* (5'-GGAGG-3') was located at the loop and partially blocked in the stem of the structure. Thus, even though the sequence of the RBS of *gutM* indicates a high efficiency of utilization for the ribosome, the initiation of translation of *gutM* could be partially impaired by the partial RBS blockage, which would be released by the opening provoked by the passage of the ribosomes translating *gutR*. In addition, the overlapping of *gutR* and *gutM* is located at the end of the stem of the structure. Thus, two post-transcriptional regulations could take place: (i) translation of *gutR* can act by favoring translation of *gutM* by exposition of its RBS and (ii) a -1 frameshift (Atkins et al., 2016) could happen at the TAA termination codon of *gutR* and ribosomes translating this could step back one nucleotide and upon charging the corresponding tRNA read the Leu (TTA) codon and continue translating *gutM*. In this way a fused peptide GutR-M could be synthesized. The same structure could be formed in the transcript encoded by the plasmid pRCR19 with a  $\Delta G = -5.9 \text{ kcal mol}^{-1}$ , containing *gutRM* and which could be a substrate for the two proposed post-transcriptional regulations. Furthermore, the DNA fragment cloned in pRCR18, lacks most of the *gutR* gene but still retains some of the 3'-end region of this gene and the encoded mRNA can form a secondary structure almost identical to the wild-type structure (with only a change of A-U by G-U pairing at the end of the stem, Figure 9). Thus in bacteria carrying pRCR18 partial blockage of the RBS could take place, but synthesis of GutR-M could not occur. This could explain the antagonistic effect of overexpression of *gutM* from pRCR18 (increase of expression from  $P_{gut}$ ) and pRCR19 (decrease of expression from  $P_{gut}$ ), if GutR-M exists and has a role.

Prediction of transmembrane regions in the regulatory proteins with the TM-Pred revealed that GutR is a soluble protein and that the first amino acids from 1 to 21 of GutM constituted a transmembrane region also predicted for the GutR-M fused polypeptide. This fused polypeptide could provide an efficient anchoring of the regulator to the membrane bringing it close to the  $PTS^{gut}$  system facilitating the phosphorylation of GutR and resulting in the physiological optimal expression of the operon. This generation of a fused polypeptide could also take place in *L. plantarum* but does not seem to occur in *L. casei*, since in this bacterium the TAA translational stop codon of GutR and the ATG start codon of GutM are adjacent and not overlapped (Figure 9). However, the *L. casei* *gut* transcript can form a secondary structure with a  $\Delta G = -9.8 \text{ kcal mol}^{-1}$  which could block the RBS of *gutM* gene, couple translation of GutR and GutM could take place, and protein-protein interaction could be responsible for higher activation of the system at the beginning of the induction process. Our results indicate that high levels of GutM synthesized from a multicopy plasmid have a deleterious effect for the bacteria (Figure 8) and probably the proposed models of posttranscriptional regulation are designed to have the right concentration of regulatory proteins. Nevertheless, further experiments are required to pinpoint the role of GutM and of the putative GutR-M polypeptide of *P. parvulus*.

## AUTHOR CONTRIBUTIONS

AP-R contributed to all parts of the experimental work and wrote a draft of the manuscript. MW performed the initial detection

of sorbitol utilization and characterization of *gut* genes. AP contributed to the characterization of the sorbitol metabolism. PR participated in the elaboration of the manuscript and analysis of the DNA sequences. GS contributed to the design and analysis of the experimental work involving characterization of regulation of *gut* operon expression. MM contributed to the design of strategies to determine trans complementation of the *gut* operon and corrected the manuscript. PL participated in study conception, data interpretation and generated the final version of the manuscript. All authors have read and approved the final manuscript.

## FUNDING

This work was supported by the Spanish Ministry of Economy and Competitiveness (grant AGL2015-65010-C3-1-R).

## ACKNOWLEDGMENTS

The authors thank Dr. Stephen W. Elson for critical reading of the manuscript and to Dr. M<sup>a</sup> Teresa Dueñas for providing the original *P. parvulus* 2.6NR isolate to perform this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02393/full#supplementary-material>

## REFERENCES

- Alcantara, C., Sarmiento-Rubiano, L. A., Monedero, V., Deutscher, J., Perez-Martinez, G., and Yebra, M. J. (2008). Regulation of *Lactobacillus casei* sorbitol utilization genes requires DNA-binding transcriptional activator GutR and the conserved protein GutM. *Appl. Environ. Microbiol.* 74, 5731–5740. doi: 10.1128/AEM.00230-08
- Aldridge, P., Metzger, M., and Geider, K. (1997). Genetics of sorbitol metabolism in *Erwinia amylovora* and its influence on bacterial virulence. *Mol. Gen. Genet.* 256, 611–619.
- Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., and Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* 44, 7007–7078. doi: 10.1093/nar/gkw530
- Berthier, F., Zagorec, M., Champomier-Vergès, M., Ehrlich, S. D., and Morel-Deville, F. (1996). Efficient transformation of *Lactobacillus sake* by electroporation. *Microbiology* 142, 1273–1279. doi: 10.1099/13500872-142-5-1273
- Boyd, D. A., Thevenot, T., Gumbmann, M., Honeyman, A. L., and Hamilton, I. R. (2000). Identification of the operon for the sorbitol (glucitol) phosphoenolpyruvate: sugar phosphotransferase system in *Streptococcus mutans*. *Infect. Immun.* 68, 925–930.
- Claesson, M. J., Li, Y., Leahy, S., Canchaya, C., van Pijkeren, J. P., Cerdeño-Tárraga, A. M., et al. (2006). Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6718–6723. doi: 10.1073/pnas.0511060103
- Cui, Y., Hu, T., Qu, X., Zhang, L., Ding, Z., Dong, A., et al. (2015). Plasmids from food lactic acid bacteria: diversity, similarity, and new developments. *Int. J. Mol. Sci.* 16, 13172–13202.
- de Man, J. C., Rogosa, M., and Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J. Appl. Microbiol.* 23, 130–135. doi: 10.1111/j.1365-2672.1960.tb00188.x
- Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* 11, 87–93. doi: 10.1016/j.mib.2008.02.007
- Dornan, S., and Collins, M. (1987). High efficiency electroporation of *Lactococcus lactis* subsp. *lactis* LM0230. *Lett. Appl. Microbiol.* 11, 62–64. doi: 10.1111/j.1472-765X.1990.tb01275.x
- Dueñas-Chasco, M. T., Rodríguez-Carvajal, M. A., Mateo, P. T., Franco-Rodríguez, G., Espartero, J. L., Irastorza-Iribas, A., et al. (1997). Structural analysis of the exopolysaccharide produced by *Pediococcus damnosus* 2.6. *Carbohydr. Res.* 303, 453–458. doi: 10.1016/S0008-6215(97)00192-4
- Favier, M., Bihère, E., Lonvaud-Funel, A., Moine, V., and Lucas, P. M. (2012). Identification of pOENI-1 and related plasmids in *Oenococcus oeni* strains performing the malolactic fermentation in wine. *PLOS ONE* 7:e49082. doi: 10.1371/journal.pone.0049082
- Fernández-Bañares, F., Esteve, M., and Viver, J. M. (2009). Fructose-sorbitol malabsorption. *Curr. Gastroenterol. Rep.* 11, 368–374.
- Fernández, K., Dueñas, M. T., Irastorza, A., Bilbao, A., and del Campo, G. (1995). Characterization and DNA plasmid analysis of ropy *Pediococcus* spp. strains isolated from Basque Country ciders. *J. Food Prot.* 59, 35–40.
- Fernández de Palencia, P., Werning, M. L., Sierra-Filardi, E., Dueñas, M. T., Irastorza, A., Corbí, A. L., et al. (2009). Probiotic properties of the 2-substituted (1,3)- $\beta$ -D-glucan-producing bacterium *Pediococcus parvulus* 2.6. *Appl. Environ. Microbiol.* 75, 4887–4891. doi: 10.1128/AEM.00394-09
- Immerstrand, T., Paul, C. J., Rosenquist, A., Deraz, S., Mårtensson, O. B., Ljungh, A., et al. (2010). Characterization of the properties of *Pediococcus parvulus* for probiotic or protective culture use. *J. Food Prot.* 73, 960–966.



- Jiménez, E., Martín, R., Maldonado, A., Martín, V., Gómez de Segura, A., Fernández, L., et al. (2010). Complete genome sequence of *Lactobacillus salivarius* CECT 5713, a probiotic strain isolated from human milk and infant feces. *J. Bacteriol.* 192, 5266–5267. doi: 10.1128/jb.00703-10
- Ladero, V., Ramos, A., Wiersma, A., Goffin, P., Schanck, A., Kleerebezem, M., et al. (2007). High-level production of the low-calorie sugar sorbitol by *Lactobacillus plantarum* through metabolic engineering. *Appl. Environ. Microbiol.* 73, 1864–1872. doi: 10.1128/AEM.02304-06
- Lamontanara, A., Caggianiello, G., Orrù, L., Capozzi, V., Michelotti, V., Bayjanov, J. R., et al. (2015). Draft genome sequence of *Lactobacillus plantarum* Lp90 isolated from wine. *Genome Announc.* 3:e00097-15. doi: 10.1128/genomeA.00097-15
- López, P., Martínez, S., Diaz, A., Espinosa, M., and Lacks, S. A. (1989). Characterization of the *polA* gene of *Streptococcus pneumoniae* and comparison of the DNA polymerase I it encodes to homologous enzymes from *Escherichia coli* and phage T7. *J. Biol. Chem.* 264, 4255–4263.
- Macrina, F. L., Kopecko, D. J., Jones, K. R., Ayers, D. J., and McCowen, S. M. (1978). A multiple plasmid-containing *Escherichia coli* strain: convenient source of size reference plasmid molecules. *Plasmid* 1, 417–420. doi: 10.1016/0147-619X(78)90056-2
- Mazé, A., Boël, G., Zúñiga, M., Bourand, A., Loux, V., Yebra, M. J., et al. (2010). Complete genome sequence of the probiotic *Lactobacillus casei* strain BL23. *J. Bacteriol.* 192, 2647–2648. doi: 10.1128/JB.00076-10
- Mohedano, M. L., García-Cayuela, T., Pérez-Ramos, A., Gaiser, R., Requena, T., and López, P. (2015). Construction and validation of a mCherry protein vector for promoter analysis in *Lactobacillus acidophilus*. *J. Ind. Microbiol. Biotechnol.* 42, 247–253. doi: 10.1007/s10295-014-1567-4
- Montalto, M., Gallo, A., Ojetti, V., and Gasbarrini, A. (2013). Fructose, trehalose and sorbitol malabsorption. *Eur. Rev. Med. Pharmacol. Sci.* 17, 26–29.
- Nissen, L., Perez-Martinez, G., and Yebra, M. J. (2005). Sorbitol synthesis by an engineered *Lactobacillus casei* strain expressing a sorbitol-6-phosphate dehydrogenase gene within the lactose operon. *FEMS Microbiol. Lett.* 249, 177–183. doi: 10.1016/j.femsle.2005.06.010
- Notararigo, S., Casas-Engel, M., Fernández de Palencia, P., Corbí, A. L., and López, P. (2014). Immunomodulation of human macrophages and myeloid cells by 2-substituted (1-3)- $\beta$ -D-glucan from *P. Parvulus* 2.6. *Carbohydr. Polym.* 112, 109–113. doi: 10.1016/j.carbpol.2014.05.073
- Notararigo, S., Nacher-Vázquez, M., Ibarburu, I., Werning, M. L., de Palencia, P. F., Duenas, M. T., et al. (2013). Comparative analysis of production and purification of homo- and hetero-polysaccharides produced by lactic acid bacteria. *Carbohydr. Polym.* 93, 57–64. doi: 10.1016/j.carbpol.2012.05.016
- Pérez-Ramos, A., Mohedano, M. L., Puertas, A., Lamontanara, A., Orru, L., Spano, G., et al. (2016). Draft genome sequence of *Pediococcus parvulus* 2.6, a probiotic  $\beta$ -glucan producer strain. *Genome Announc.* 4:e01381-16. doi: 10.1128/genomeA.01381-16
- Raftis, E. J., Forde, B. M., Claesson, M. J., and O'Toole, P. W. (2014). Unusual genome complexity in *Lactobacillus salivarius* JCM1046. *BMC Genomics* 15:771. doi: 10.1186/1471-2164-15-771
- Rhodes, M. W., and Kator, H. (1999). Sorbitol-fermenting bifidobacteria as indicators of diffuse human faecal pollution in estuarine watersheds. *J. Appl. Microbiol.* 87, 528–535.
- Russo, P., Iturria, I., Mohedano, M. L., Caggianiello, G., Rainieri, S., Fiocco, D., et al. (2015). Zebrafish gut colonization by mCherry-labelled lactic acid bacteria. *Appl. Microbiol. Biotechnol.* 99, 3479–3490. doi: 10.1007/s00253-014-6351-x
- Russo, P., López, P., Capozzi, V., Fernández de Palencia, P., Dueñas, M. T., Spano, G., et al. (2012). Beta-glucans improve growth, viability and colonization of probiotic microorganisms. *Int. J. Mol. Sci.* 13, 6026–6039. doi: 10.3390/ijms13056026
- Sarmiento-Rubiano, L. A., Zuniga, M., Perez-Martinez, G., and Yebra, M. J. (2007). Dietary supplementation with sorbitol results in selective enrichment of lactobacilli in rat intestine. *Res. Microbiol.* 158, 694–701. doi: 10.1016/j.resmic.2007.07.007
- Silveira, M. M., and Jonas, R. (2002). The biotechnological production of sorbitol. *Appl. Microbiol. Biotechnol.* 59, 400–408. doi: 10.1007/s00253-002-1046-0
- Stülke, J., Arnaud, M., Rapaport, G., and Martin-Verstraete, I. (1998). PRD — a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Mol. Microbiol.* 28, 865–874.
- Tangney, M., Brehm, J. K., Minton, N. P., and Mitchell, W. J. (1998). A gene system for glucitol transport and metabolism in *Clostridium beijerinckii* NCIMB 8052. *Appl. Environ. Microbiol.* 64, 1612–1619.
- Velasco, S. E., Areizaga, J., Irastorza, A., Dueñas, M. T., Santamaria, A., and Muñoz, M. E. (2009). Chemical and rheological properties of the  $\beta$ -glucan produced by *Pediococcus parvulus* 2.6. *J. Agric. Food Chem.* 57, 1827–1834. doi: 10.1021/jf803065w
- Velasco, S. E., Yebra, M. J., Monedero, V., Ibarburu, I., Dueñas, M. T., and Irastorza, A. (2007). Influence of the carbohydrate source on  $\beta$ -glucan production and enzyme activities involved in sugar metabolism in *Pediococcus parvulus* 2.6. *Int. J. Food Microbiol.* 115, 325–334. doi: 10.1016/j.ijfoodmicro.2006.12.023
- Wegmann, U., O'connell-Motherway, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., et al. (2007). Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J. Bacteriol.* 189, 3256–3270. doi: 10.1128/JB.01768-06
- Werning, M. L., Ibarburu, I., Dueñas, M. T., Irastorza, A., Navas, J., and López, P. (2006). *Pediococcus parvulus* *gtf* gene encoding the GTF glycosyltransferase and its application for specific PCR detection of beta-D-glucan-producing bacteria in foods and beverages. *J. Food Prot.* 69, 161–169.
- Werning, M. L., Pérez-Ramos, A., Fernández de Palencia, P., Mohedano, M. L., Dueñas, M. T., Prieto, A., et al. (2014). A specific immunological method to detect and quantify bacterial 2-substituted (1,3)- $\beta$ -D-glucan. *Carbohydr. Polym.* 113, 39–45. doi: 10.1016/j.carbpol.2014.06.072
- Yamada, M., and Saier, M. H. J. (1988). Positive and negative regulators for glucitol (gut) operon expression in *Escherichia coli*. *J. Mol. Biol.* 203, 569–583.
- Yebra, M. J., and Pérez-Martínez, G. (2002). Cross-talk between the L-sorbose and D-sorbitol (D-glucitol) metabolic pathways in *Lactobacillus casei*. *Microbiology* 148, 2351–2359.
- Zarour, K., Vieco, N., Pérez-Ramos, A., Nacher-Vázquez, M., Mohedano, M. L., and López, P. (2017). “Food ingredients synthesised by lactic acid bacteria,” in *Microbial Production of Ingredients and Additives*, eds A. M. Holban and A. M. Grumezescu (Cambridge: Elsevier), 89–124.
- Zumbé, A., Lee, A., and Storey, D. (2001). Polyols in confectionery: the route to sugar-free, reduced sugar and reduced calorie confectionery. *Br. J. Nutr.* 85, S31–S45.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pérez-Ramos, Werning, Prieto, Russo, Spano, Mohedano and López. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Dextranucrase Expression Is Concomitant with that of Replication and Maintenance Functions of the pMN1 Plasmid in *Lactobacillus sakei* MN1

Montserrat Nácher-Vázquez<sup>1</sup>, José A. Ruiz-Masó<sup>1</sup>, María L. Mohedano<sup>1</sup>, Gloria del Solar<sup>1</sup>, Rosa Aznar<sup>2,3</sup> and Paloma López<sup>1\*</sup>

<sup>1</sup> Department of Molecular Microbiology and Infection Biology, Biological Research Center, Spanish National Research Council (CSIC), Madrid, Spain, <sup>2</sup> Department of Food Safety and Preservation, Institute of Agrochemistry and Food Technology, CSIC, Paterna, Spain, <sup>3</sup> Department of Microbiology and Ecology, University of Valencia, Burjassot, Spain

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Preeti Srivastava,  
Indian Institute of Technology Delhi,  
India

Stephen M. Kwong,  
University of Sydney, Australia  
Reinaldo Fraga,

Cuban Research Institute on  
Sugarcane By-Products (ICIDCA),  
Cuba

### \*Correspondence:

Paloma López  
plg@cib.csic.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 August 2017

**Accepted:** 06 November 2017

**Published:** 21 November 2017

### Citation:

Nácher-Vázquez M, Ruiz-Masó JA, Mohedano ML, del Solar G, Aznar R and López P (2017) Dextranucrase Expression Is Concomitant with that of Replication and Maintenance Functions of the pMN1 Plasmid in *Lactobacillus sakei* MN1. *Front. Microbiol.* 8:2281. doi: 10.3389/fmicb.2017.02281

The exopolysaccharide synthesized by *Lactobacillus sakei* MN1 is a dextran with antiviral and immunomodulatory properties of potential utility in aquaculture. In this work we have investigated the genetic basis of dextran production by this bacterium. Southern blot hybridization experiments demonstrated the plasmidic location of the *dsrLS* gene, which encodes the dextranucrase involved in dextran synthesis. DNA sequencing of the 11,126 kbp plasmid (pMN1) revealed that it belongs to a family which replicates by the theta mechanism, whose prototype is pUCL287. The plasmid comprises the origin of replication, *repA*, *repB*, and *dsrLS* genes, as well as seven open reading frames of uncharacterized function. *Lb. sakei* MN1 produces dextran when sucrose, but not glucose, is present in the growth medium. Therefore, plasmid copy number and stability, as well as *dsrLS* expression, were investigated in cultures grown in the presence of either sucrose or glucose. The results revealed that pMN1 is a stable low-copy-number plasmid in both conditions. Gene expression studies showed that *dsrLS* is constitutively expressed, irrespective of the carbon source present in the medium. Moreover, *dsrLS* is expressed from a monocistronic transcript as well as from a polycistronic *repA-repB-orf1-dsrLS* mRNA. To our knowledge, this is the first report of a plasmid-borne dextranucrase-encoding gene, as well as the first time that co-transcription of genes involved in plasmid maintenance and replication with a gene encoding an enzyme has been established.

**Keywords:** dextran, dextranucrase, lactic acid bacteria, *Lactobacillus sakei*, plasmid, probiotics

## INTRODUCTION

Lactic acid bacteria (LAB) play an important role in the production of fermented foods based on milk, meat, and vegetables as well as alcoholic beverages, due to their metabolic pathways, whose products contribute to food safety (e.g., lactic acid or hydrogen peroxide) and to the organoleptic characteristics [e.g., the diacetyl and other aroma compounds or texturizing exopolysaccharides (EPS)]. In addition, some LAB have beneficial

health characteristics (probiotic properties) or metabolic capacities such as the production of enzymes (amylases, phytases), vitamins (folates, riboflavin), or EPS, which are of particular interest for the agro-food industry and for the formulation of new functional foods (Anastasio et al., 2010; Badel et al., 2011; Capozzi et al., 2012).

The EPS produced by LAB can be classified by (i) composition, because they include different types of bonds and monosaccharide subunits; (ii) types and degrees of branching; (iii) molecular mass; and (iv) three-dimensional (3D) structural conformation. EPS are classified as homopolysaccharides (HoPS), consisting of a single type of monosaccharide, or heteropolysaccharides (HePS), composed of two or more types of monosaccharides. HoPS are glucans, fructans, or galactans, made up of repeating units of glucose, fructose, or galactose, respectively (Pérez-Ramos et al., 2015).

The  $\alpha$ -D-glucans are the most widely produced HoPS and according to the linkage in the main chain, they are subdivided in dextrans  $\alpha$ -(1,6), mutans  $\alpha$ -(1,3), reuterans  $\alpha$ -(1,4), and alternans  $\alpha$ -(1,3) and  $\alpha$ -(1,6) and may present different types and degrees of branching (Monsan et al., 2001). Of these, dextrans are currently used in the food and pharmaceutical industries (Aman et al., 2012). The viscosity and rheological properties of the dextran solutions are influenced by their molecular masses and consequently define their applications. Thus, low-molecular mass dextrans are used in the photographic and pharmaceutical industries, whereas the high-molecular mass dextrans are utilized in the chemical industry. Among various medical applications, dextrans are used for anticoagulant therapy as heparin substitutes and blood plasma replacers/expanders. In addition, there is evidence that dextran sulfate has an antiviral effect against human immunodeficiency virus (Piret et al., 2000) and we have recently demonstrated that dextrans synthesized by LAB have potential as antivirals and immunomodulatory agents in trout (Nácher-Vázquez et al., 2015). In the food industry HoPS are added to bakery products and confectionery to improve softness or moisture retention, to prevent crystallization, and to increase viscosity, rheology, texture, and volume (Pérez-Ramos et al., 2015). They are also used as films to protect surfaces of frozen fish, meat, vegetables, or cheese from oxidation and other chemical changes.

Dextrans are synthesized by dextranucleases (Dsr), which catalyze the transfer of D-glucopyranosyl residues from sucrose to the growing polymer, accompanied by fructose release (Werning et al., 2012). The Dsr-encoding genes are carried by strains belonging to the genera *Lactobacillus*, *Leuconostoc*, *Oenococcus*, *Streptococcus*, *Weissella*, and *Pediococcus* (Kralj et al., 2004a; Naessens et al., 2005; Bounaix et al., 2010; Werning et al., 2012; Amari et al., 2013; Rühmkorf et al., 2013; Dimopoulou et al., 2015; Yanping et al., 2015). Many Dsr have been characterized, but despite the interest of dextrans in various applications, little is known about the regulation of the expression of Dsr. By determining the levels or the activity of the Dsr in bacterial cultures grown in presence of sucrose or other sugars it has been inferred that their expression may be constitutive or sucrose-inducible. However, transcriptional analysis of *dsr* genes expression to validate sucrose inducibility has been performed in

a few cases only (e.g., in *Leuconostoc mesenteroides* NRRL B-512F, Quirasco et al., 1999).

We have previously demonstrated that *Lb. sakei* MN1 isolated from a fermented meat product synthesizes an  $\alpha$ -(1-6) glucan with ~6% substitution, at positions O-3, by side chains composed of a single residue of glucose and with a molecular mass of  $1.7 \times 10^8$  Da (Nácher-Vázquez et al., 2015; Zarour et al., 2017). We have performed *in vitro* and *in vivo* experiments that support that this dextran has antiviral and immunomodulatory properties of interest in aquaculture (Nácher-Vázquez et al., 2015). Moreover, we have demonstrated that the purified HoPS is able to efficiently immunomodulate *in vitro* human macrophages (Zarour et al., 2017). In addition, we have provided evidences that *Lb. sakei* MN1 has probiotic properties and we have shown that the production of dextran influences, *in vitro*, the bacterial capability for aggregation, biofilm formation, and adhesion to enterocytes as well as *in vivo* bacterial colonization and competition with pathogens in gnotobiotic zebrafish models (Nácher-Vázquez et al., 2017). Therefore this bacterium and its dextran seem to have potential for development of functional synbiotic food and feed. Thus, in this work, we have characterized the genetic basis of dextran production in this bacterium with the aim of having a better knowledge of the practical utility of *Lb. sakei* MN1.

## MATERIALS AND METHODS

### Bacterial Strains and Growth Media

The bacterial strains used in this work are shown in Table 1. The *Latococcus lactis* strains were grown in M17 broth (Oxoid) supplemented with 0.5% glucose (M17G) or 0.5% glucose plus 0.8% sucrose (M17GS) and *Lb. sakei* strains were grown in Man Rogosa Sharpe broth (de Man et al., 1960) with 2% glucose (MRSG) or 2% sucrose (MRSS) or in defined medium with 0.8% glucose (CDMG) or 0.8% sucrose (CDMS) (Sánchez et al., 2008) and incubated at 30°C. *Escherichia coli* strains were grown in LB medium containing 10 g L<sup>-1</sup> of tryptone, 5 g L<sup>-1</sup> of yeast extract, and 10 g L<sup>-1</sup> of NaCl (pH 7.0) and incubated at 37°C. When the bacteria carried pPCR or pPCR-based plasmid derivatives conferring resistance to chloramphenicol (Cm) the medium was supplemented with Cm at 5 µg mL<sup>-1</sup> for *L. lactis* and *Lb. sakei* strains or at 10 µg mL<sup>-1</sup> for *E. coli* DH5 $\alpha$ .

### Genomic and Plasmidic DNA Preparations

To isolate plasmidic DNA, bacterial cultures were grown to an absorbance at 600 nm (A<sub>600</sub>) of 2 at 30°C in MRSG and 10 mL of each culture were sedimented by centrifugation (15,700  $\times$  g, 10 min, 4°C). For cellular lysis, cells were washed with phosphate buffer saline (PBS, pH 7.4), resuspended in 2 mL of a solution containing 25% sucrose, 30 mg mL<sup>-1</sup> lysozyme, 120 U mL<sup>-1</sup> mutanolysin, and 40 µg mL<sup>-1</sup> RNase A, and incubated for 15 min at 37°C. Then, cell debris and chromosomal DNA were removed from the extracts by: (i) treatment for 7 min at 21°C with 4 mL of a solution containing 0.13 N NaOH and 2% SDS, (ii) incubation for 15 min at 0°C with 3 mL of 1 M potassium acetate pH 4.8, and (iii) sedimentation (centrifugation at 15,700  $\times$  g, 15 min, 4°C). The plasmidic DNA present in the supernatants was precipitated and concentrated by

**TABLE 1** | Description of bacteria used in this work.

Strain	Plasmids	<sup>a</sup> Resistance	Information	References
<i>Lb. sakei</i> MN1 (CECT 8329)	pMN1 pMN2	–	Isolated from a fermented meat product	Chenoll et al., 2007
<i>Lb. sakei</i> MN1[pRCR12]	pMN1 pMN2 pRCR12	Cm <sup>R</sup>	Derivative of pRCR. It contains the promoter Px of <i>malXCD</i> operon of <i>Streptococcus pneumoniae</i> JNR7/87 fused to the <i>mrfp</i> gene	Nácher-Vázquez et al., 2017
<i>Lb. sakei</i> MN1[pRCR13]	pMN1 pMN2 pRCR13	Cm <sup>R</sup>	Derivative of pRCR by cloning of region A (GenBank MF590088 position: 38–363) upstream of <i>mrfp</i>	This study
<i>Lb. sakei</i> MN1[pRCR14]	pMN1 pMN2 pRCR14	Cm <sup>R</sup>	Derivative of pRCR by cloning of region B (GenBank MF590088 position: 1,476–1,819) upstream of <i>mrfp</i>	This study
<i>Lb. sakei</i> MN1[pRCR15]	pMN1 pMN2 pRCR15	Cm <sup>R</sup>	Derivative of pRCR by cloning of region C (GenBank MF590088 position: 2,392–2,749) upstream of <i>mrfp</i>	This study
<i>L. lactis</i> MG1363	–	–	Strain derived from <i>L. lactis</i> 712 by plasmids curing	Gasson, 1983
<i>L. lactis</i> MG1363[pRCR13]	pRCR13	Cm <sup>R</sup>	Derivative of pRCR by cloning of region A upstream of <i>mrfp</i>	This study
<i>L. lactis</i> MG1363[pRCR14]	pRCR14	Cm <sup>R</sup>	Derivative of pRCR by cloning of region B upstream of <i>mrfp</i>	This study
<i>L. lactis</i> MG1363[pRCR15]	pRCR15	Cm <sup>R</sup>	Derivative of pRCR by cloning of region C upstream of <i>mrfp</i>	This study
<i>E. coli</i> DH5α[pRCR]	pRCR	Cm <sup>R</sup>	Promoters probe vector containing the <i>mrfp</i> gene, which encodes the fluorescent mCherry protein	Mohedano et al., 2015
<i>E. coli</i> V517	Eight plasmids pVA517A through pVA517H	ND	Source of plasmids used as references in agarose gel analysis	Macrina et al., 1978

<sup>a</sup>ND, No determined; Cm<sup>R</sup>, resistance to chloramphenicol.

addition of 42% (final concentration) isopropanol, centrifugation as above and dissolution in 3.2 mL of ultrapure water. The DNA preparation was purified and deproteinated by treatment with 2 mL of a solution containing 7.5 M ammonium acetate and ethidium bromide at 0.5 mg mL<sup>-1</sup>, addition of 1:1 (v/v) of a mixture of phenol, chloroform, and isoamyl alcohol (50:48:2, vol/vol/vol) and centrifugation (15,700 × g, 10 min, 21°C). Plasmidic DNA was precipitated from the aqueous phase with 69.5% (final concentration) ethanol for 12 h at –20°C and recovered by centrifugation (11,269 × g, 45 min, –10°C). The precipitated DNA was washed with 1 mL of 70% ethanol, sedimented by centrifugation (11,269 × g, 30 min, –10°C), and dissolved in 10 mM Tris buffer pH 8.0 (100 µL).

For the isolation of genomic DNA, bacterial cultures were grown to an A<sub>600</sub> = 2 at 30°C in MRSG and 1 mL of each culture were sedimented by centrifugation (15,700 × g; 10 min, 4°C). For cellular lysis, cells were washed with phosphate buffer saline (PBS, pH 7.4), resuspended in a solution (100 µL) containing 25% sucrose, 50 mM Tris pH 8.0, 0.1 M NaCl, 30 mg mL<sup>-1</sup> lysozyme, 240 U mL<sup>-1</sup> mutanolysin, and 80 µg mL<sup>-1</sup> RNase A, and incubated for 15 min at 37°C. Then, lysed cells were

treated with 1% (final concentration) SDS for 2 min at 21°C and passed through a 25 GA needle (0.5 × 16 mm) three times. The extracts were deproteinated by treatment with an equal volume of a mixture of phenol, chloroform, and isoamyl alcohol (50:48:2, vol/vol/vol) for 5 min at 21°C and centrifugation (15,700 × g, 10 min, 21°C). The DNA contained in the aqueous phase was precipitated with 69.5% (final concentration) ethanol and 83 mM (final concentration) sodium acetate pH 7.0 for 12 h at –20°C and recovered by centrifugation (11,269 × g, 45 min, –10°C). The precipitated DNA was washed with 1 mL of 70% ethanol, sedimented by centrifugation (11,269 × g, 30 min, –10°C), and dissolved in 10 mM Tris buffer pH 8.0 (100 µL).

## Total RNA Preparations

For the isolation of total RNA, bacterial cultures were grown to A<sub>600</sub> = 2 at 30°C in CDMG or CDMS for RT-PCR. Total RNAs were isolated using the kit “FastRNA Pro Blue” (QBIogene) and subjected to electrophoresis in 0.8% agarose gels at a constant voltage of 135 V for 20 min to check the integrity of the rRNAs. The total RNA concentration was determined with a Qubit 2.0 fluorimeter (Invitrogen) following the instructions of



the supplier. To ensure absence of DNA, the RNA preparations were incubated for 1 h at 37°C with 1 µg mL<sup>-1</sup> DNase I (Sigma-Aldrich) and then purified following the “RNA Cleanup” protocol of the “RNeasy Midi” kit from QIAGEN. In addition, the samples were subjected to three cycles of deproteinization, which involved incubation with acid phenol for 5 min at 70°C with shaking and subsequent centrifugation (15,700 × g, 21°C, 5 min). The collected supernatants were treated with one volume of phenol:chloroform-isoamyl alcohol (50:48:2) and centrifuged at 15,700 × g, 21°C, 10 min). The RNA present in the aqueous phase was precipitated by the addition of 1/10 volume of 3 M sodium acetate (pH 7.0) and three volumes of absolute ethanol, followed by storage at -20°C for 12 h. The precipitated RNA was recovered by centrifugation (11,269 × g, -10°C, 45 min), washed with 1 mL of 75% ethanol and finally dissolved in 200 µL of diethylpyrocarbonate-treated water.

## Detection of the Genes

### Oligonucleotides, PCR, RT-PCR, and Sequencing

The oligonucleotides used are summarized in Table 2. The nucleotide sequences of 14 bacterial genes encoding dextranuclease enzymes, obtained from GenBank of the National Centre for Biotechnology (NCBI, USA), were analyzed with the BLAST program in order to design the primers *dsrF* and *dsrR* for further amplification of a conserved dextranuclease coding region, located at the catalytic site of these enzymes. Genomic DNA and plasmidic DNA of *Lb. sakei* MN1 were used as templates in a reaction with primers *dsrF* and *dsrR* with Phusion Hot Start High Fidelity Polymerase (HSHP, ThermoFisher Scientific) following the instructions of the enzyme supplier.

For RT-PCR, total RNA preparations were used. For the synthesis of the cDNAs, 400 ng of RNA, 0.75 µM of each of the oligonucleotides complementary to different regions of pMN1 and 500 µM of each of the 4 dNTPs (dATP, dGTP, dCTP, and dTTP) were used. The mixtures were incubated for 5 min at 65°C, transferred to 4°C and to each were added 1X cDNA synthesis buffer (Invitrogen), 5 mM DTT (Invitrogen), 2 U µL<sup>-1</sup> of RNaseOUT (Invitrogen), and 0.75 U µL<sup>-1</sup> of avian reverse transcriptase from ThermoScript (Invitrogen). The reaction was allowed to proceed for 1 h at 50°C and samples were next incubated 5 min at 85°C to stop the reaction. To remove RNA residues, 0.2 µg µL<sup>-1</sup> of RNase A (Sigma-Aldrich) was added and the mixture was incubated for 20 min at 37°C. Finally, the samples were dialyzed for 30 min against 10 mM Tris pH 8.0 using membranes of the Millipore V series (Merck). PCRs were carried out according to the instructions of the Thermo Fisher Scientific Phusion DNA polymerase using the primers described in Table 2.

DNA sequencing was performed by the dideoxynucleotide method at Secugen (Madrid, Spain). For the determination of the nucleotide sequence of the pMN1 plasmid, the walking strategy was followed, after detection of the *dsrLS* gene by PCR with the *dsrF* and *dsrR* primers and further sequencing of the amplicon. To determine the sequence of pMN1 by the dideoxynucleotide method, it was used as substrate the product of a polymerization reaction catalyzed by the bacteriophage Φ29 DNA polymerase

with plasmidic DNA of *Lb. sakei* MN1 and hexamers containing random sequences. The DNA sequence of plasmid pMN1 was deposited in GenBank (accession No MF590088).

## Southern Blot Hybridization

Genomic and plasmidic DNA preparations were subjected to electrophoresis in a 0.7% agarose gel with 40 mM Tris, 20 mM acetate, and 1 mM EDTA buffer at constant amperage of 40 mA for ~4 h, and DNA molecules were revealed by staining with ethidium bromide at 0.5 µg mL<sup>-1</sup>. Then, DNA was transferred to a 0.45 µm nylon membrane (Biodyne A, Pall Corporation) and hybridized with the probe. The temperature of hybridization was 45°C. Probe labeling and detection procedures were performed with the NEBlot® Phototope® Kit (New England BioLabs) and the Phototope®-Stars Detection Kit (New England Biolabs). Substrate for probe labeling was the 695-bp amplicon synthesized with primers *dsrF* and *dsrR* (Table 2) and *Lb. sakei* MN1 genomic DNA following the indications of the kit supplier.

## Determination of pMN1 Plasmid Copy Number

### Preparation of Template DNA for Real Time-qPCR

Exponential cultures ( $A_{600} = 1.0$ ) of *Lb. sakei* MN1 grown in either MRSG or MRSS at 30°C were generated by inoculation of the media (dilution 1/1,000) with a bacterial stock culture previously grown in MRSG. These two cultures (designated generation 0) were always maintained in the exponential phase and sub-cultured in the corresponding medium by dilutions 1/1,000 for 60 more generations. Then genomic DNA was isolated from ~0.5 × 10<sup>9</sup> bacteria of the 0 and 60 generations cultures by using the Wizard® Genomic DNA Purification Kit (Promega). At the cell lysis step, 30 mg mL<sup>-1</sup> of lysozyme and 30 U of mutanolysin were added. Concentration of the genomic DNA was determined with a Qubit fluorometer by using the Qubit HS dsDNA Assay Kit (Molecular Probes). The DNA extracts were digested with EcoRI, a restriction enzyme that linearizes the pMN1 plasmid leaving intact the *repA* and *pcrA* amplicons. This method was developed to obtain accurate qPCR-based plasmid copy numbers (Providenti et al., 2006).

## Real Time-qPCR Analysis

Two primer sets were designed, based on the pMN1 sequence (this work) and on the *Lb. sakei* 23K (NC\_007576.1) *pcrA* sequence, specific for either the pMN1 replication protein coding gene (*repA*) or the PcrA helicase chromosomal reference gene (*pcrA*). The criteria used during primer design was that the primers had a predicted T<sub>m</sub> of ~59°C and that they generated amplicons ~140-bp long.

The qPCR were conducted in a total volume of 20 µL using an iQ5 real-time detection system (BIO-RAD) and the IQ<sup>TM</sup> SYBR® Green Supermix (Bio-Rad Laboratories), following the manufacturer's recommendations. Decimally diluted EcoRI-digested total DNA preparations (15, 1.5, 0.15, and 0.015 ng per reaction) were analyzed using 0.5 µM (final concentration) of the specific forward and reverse primers. To prepare the reactions and minimize pipetting errors 2 µL of template DNA were added to individual qPCRs. Thermal cycling conditions were as follows:

**TABLE 2** | Description of primers used in this work.

Target	Primer <sup>a</sup>	Sequence (5'-3')	Product (bp)
<i>pcrA</i>	pcrF pcrR	ACAAACATGGCGCATCAACG GCGAAGGTGCTCAAGATGTTT	146
<i>repA</i>	repF repR	GGCAAGCCGTTATTGGTTAC TTTTCTGCTCTGTTAATTCACCAT	140
<i>dsrLS</i>	dsrF dsrR	GATGATGGTCAATATATGGCAA CTTGAACGATATTGGTGCCAA	695
Region A	P1F P1R	GAAGATCTTCTTTAGACCCCTCTTGAGGCT GCTCTAGAGCAGTATCATCACCTTTATCGCGC	343
Region B	P2F P2R	GGAAGATCTTCTCAGCAACAACGGTTAGCCT GCTCTAGAGCCGCCAGTGATCATATAACCGA	366
Region C	P3F P3R	GGAAGATCTTCCAAATTAACCAGAGACCGC GCTCTAGAGCTGGCTGGCTGCTAAGTAGCA	376
RT-PCR Amplicon 1	rt1F rt1R	AGCTGGGTTCGATATGCTTTA CCCACCCCTCGCTCTTTA	1,150
RT-PCR Amplicon 2	P2F rt2R	GGAAGATCTTCTCAGCAACAACGGTTAGCCT CGGTTGGCAAAGACGTTTTG	1,952
RT-PCR Amplicon 3	P3F rt2R	GGAAGATCTTCCAAATTAACCAGAGACCGC CGGTTGGCAAAGACGTTTTG	1,033
RT-PCR Amplicon 4	rt2F rt2R	ACGGCTGCGATCACTACTG CGGTTGGCAAAGACGTTTTG	496
RT-PCR Amplicon 5	rt3F rt3R	AGCTTACGCTGCTACCAAGGC ATGGCTGGAGTAAATGGATCAGCT	862

<sup>a</sup>F, Forward; R, reverse.

initial denaturation at 95°C for 5 min, followed by 40 cycles of 95°C for 10 s (denaturation), 59°C for 30 s (primer annealing), and 72°C for 20 s (elongation). A melting curve analysis of the PCR products, with a temperature gradient of 0.1°C/s from 59 to 95°C, was performed to confirm the purity and specificity of the PCR products. Two independent qPCR trials were conducted for each template source. In each trial, triplicate samples of the four different amounts of template were analyzed.

Relative copy number of pMN1 was calculated using equation (1):

$$PCN = (1 + E_{pcrA})^{Ct_{pcrA}} / (1 + E_{repA})^{Ct_{repA}}, \quad (1)$$

where  $E_{pcrA}$  and  $E_{repA}$  are, respectively, the PCR amplification efficiencies of the chromosomal and plasmidic amplicons, and  $Ct_{pcrA}$  and  $Ct_{repA}$  are the mean threshold cycle values obtained for the corresponding amplicons. A PCN value was calculated for each of the four template concentrations analyzed, and the average and standard deviation of the four values was estimated.

E values of target ( $E_{repA}$ ) and reference ( $E_{pcrA}$ ) sequences were empirically calculated for each qPCR trial. For that purpose,

mean Ct values were plotted against the logarithm of the amount of total DNA template in the assay (Figure 2C). From the slope of the curve generated by linear regression of the plotted points, the PCR amplification efficiency was determined according to the equation:

$$E = 10^{-1/\text{slope}} - 1 \quad (2)$$

Although the E values for both amplicons was higher than 0.9, we have chosen Equation (1) to calculate the relative plasmid copy number as it allows taking into account the slight differences between  $E_{\text{target}}$  and  $E_{\text{reference}}$  that we have observed.

### Construction of pRCR-pMN1 Derivative Plasmids to Detect Promoter Regions Driving Transcription of *dsrLS*

The pRCR vector containing the *mrfp* gene, which encodes the fluorescent mCherry protein, was used to detect and evaluate the performance of pMN1 promoter(s) of *dsrLS*. For the construction of pRCR13, pRCR14, and pRCR15 (pRCR derivatives, Figure S1), amplicons of 343-, 366-, and 376-bp were synthesized by PCR

with HSHFP following the instructions of the DNA polymerase supplier. The substrate for the reaction was pMN1 present in a plasmidic DNA preparation of *Lb. sakei* MN1 and the following oligonucleotide pairs, P1F and P1R for pRCR13; P2F and P2R for pRCR14; P3F and P3R for pRCR15 (Table 2). The plasmid vector pRCR, obtained from *E. coli* DH5 $\alpha$ [pRCR], and the amplicons generated by PCR were subjected to digestion with BglII and XbaI (New England Biolabs) and ligated into the pRCR vector with the T4 DNA ligase (New England Biolabs) to obtain the recombinant plasmids. The ligation mixtures were used to transform *L. lactis* MG1363 by electroporation (25  $\mu$ F, 2.5 kV and 200  $\Omega$  in 0.2 cm cuvettes), as previously described (Dornan and Collins, 1987), and transformants were selected in M17G agar plates supplemented with Cm at 5  $\mu$ g mL $^{-1}$ . The three new plasmid constructs were confirmed by automated sequencing. DNA preparations of pRCR13, pRCR14, and pRCR15 obtained from *L. lactis* MG1363 were then used to transform *Lb. sakei* MN1 by electroporation (25  $\mu$ F, 1.8 kV and 600  $\Omega$  in 0.2 cm cuvettes) as previously described (Berthier et al., 1996) and transformants were selected in MRSG-agar plates supplemented with Cm at 5  $\mu$ g mL $^{-1}$ .

### Detection of mCherry Fluorescence in LAB Carrying pRCR12, pRCR13, pRCR14, or pRCR15

To detect the expression of the mCherry fluorescent protein, *L. lactis* strains containing pRCR12, pRCR13, pRCR14, or pRCR15 were grown in 10 mL of M17G or M17GS till the initial stationary phase ( $A_{660} = 2.5$ – $2.6$ ) or to late stationary phase ( $A_{660} = 3.0$ – $3.2$ ). *Lb. sakei* strains containing the same plasmid constructs were grown in 10 mL of MRSG or MRSS medium to middle exponential phase ( $A_{600} = 2$ ) or to late exponential phase ( $A_{600} = 5$  or  $A_{600} = 10$  for cultures grown in MRSG or in MRSS, respectively). All cultures were centrifuged ( $16,000 \times g$ , 15 min, 4°C), resuspended in PBS buffer (pH 7.4), sedimented by centrifugation as above and finally resuspended in 500  $\mu$ L of PBS buffer (pH 7.4). Suspensions (200  $\mu$ L of each) were used to measure the fluorescence levels of the mCherry protein in a Varioskan Flash (Thermo Fisher Scientific) equipment, using 587 and 610 nm wavelengths for excitation and emission detection, respectively. In addition, appropriate dilutions were prepared to estimate culture biomass by measuring the absorbance at  $A_{600}$  or  $A_{480}$  for *L. lactis* or *Lb. sakei*, respectively. Three independent trials were performed and the same fresh suspensions (8  $\mu$ L of each), without fixing, were used for phase contrast and fluorescent microscopy analysis as previously described (Nácher-Vázquez et al., 2017). A Leica AF6000LX-DMI6000B model microscope (Leica Microsystems, Mannheim, Germany) was used. Illumination was provided with a 100 $\times$  objective. For detection of mCherry BP 620/60 excitation and BP 700/75 emission filters were used. Image analysis was performed using LAS AF CoreSoftware (Leica Microsystems).

### Bioinformatic Analysis of DNA Sequences and Modeling of Dsr

The DNA sequence of plasmid pMN1 was analyzed with the programs included in the DNASTAR Lasergene 12 (DNASTAR

Inc.). Homologies of pMN1 DNA sequences and of its inferred translated products with the NCBI data bases of the National Center for Biotechnology Information (NCBI) were analyzed with the Basic Local Alignment Search Tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Multiple sequence alignment of genes and proteins were performed with the Megalign (DNASTAR laser gene 12) and Clustalx 2.1 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) programs.

The primers for the qPCR experiments were designed with Primer3 v0.4.0 (Koressaar and Remm, 2007; Untergasser et al., 2012).

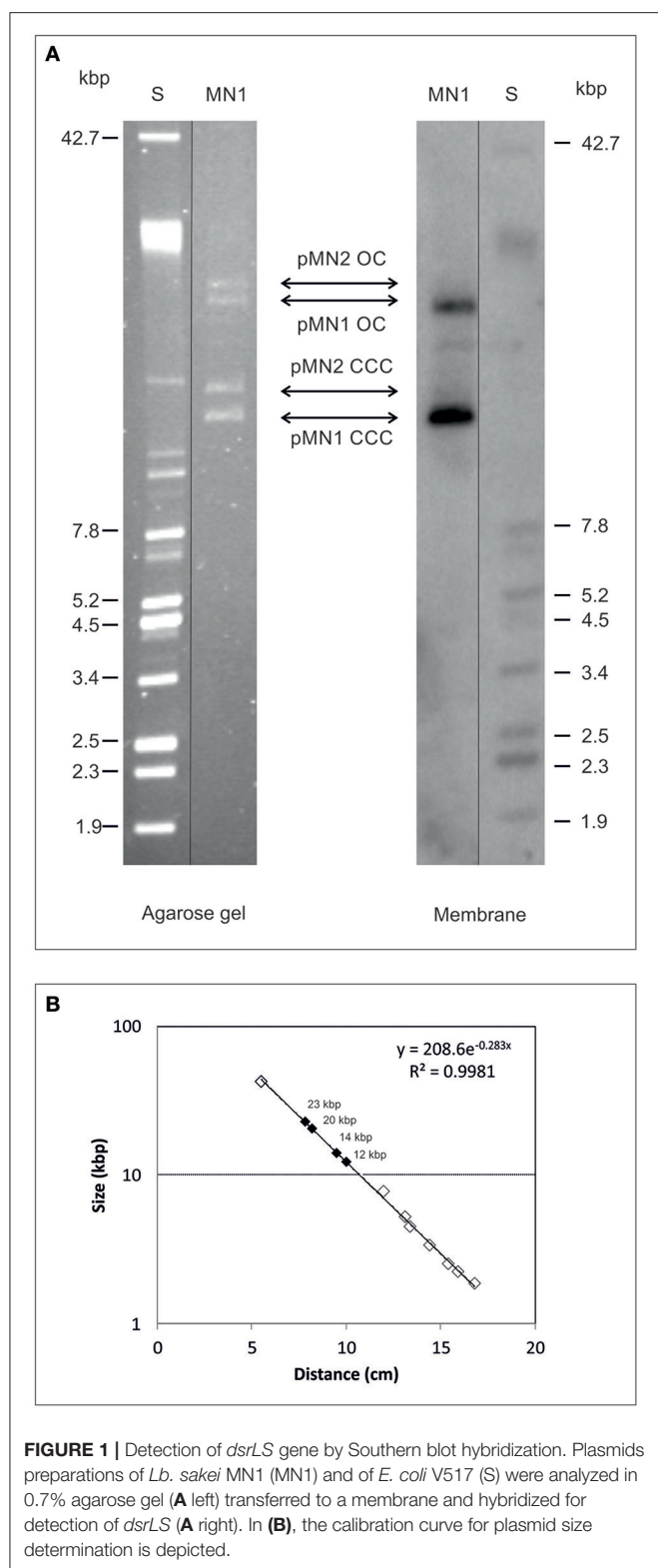
The modeling of the DsrLS was generated based on its homology with the glucanucrase of *Lactobacillus reuteri* 180, using the 3D-structure of the GTF180- $\Delta$ N glucanucrase and the I-TASSER program (Roy et al., 2010). The superposition of the DsrLS model and the 3D-structure of GTF180- $\Delta$ N was performed with the CE program at the <http://source.rcsb.org>.

## RESULTS

### Detection and Genomic Localization of *dsrLS*

Previous characterization of the EPS produced by *Lb. sakei* MN1 revealed that it is a dextran (Nácher-Vázquez et al., 2015) and indicated that this bacterium produces a Dsr (named DsrLS) responsible for the polymer synthesis. Thus, based on the known sequences of *dsr* genes from other bacteria, primers *dsrF* and *dsrR* were designed to amplify by PCR a DNA fragment located at the coding sequence of the dextranucrase catalytic domain. The expected amplicon (695-bp) was obtained using either genomic or plasmidic DNA preparations from *Lb. sakei* MN1 as template (data not shown). The determination of the nucleotide sequence of this amplicon (GenBank accession No KJ161305) and its BLAST analysis against the nucleotide (nr/nt) data base of NCBI revealed only high homologies (100, 99, and 71% identity, with 40 gaps) with genes encoding dextranucleases of *Lactobacillus curvatus* TMW1624 (GenBank accession No HE972512), *Lb. sakei* Kg15 (GenBank accession No AY697434) and *Weissella confusa* strain Cab3 (GenBank accession No KP729387.1), respectively. The overall results supported that the *dsrLS* gene has been detected, and, generation of the expected amplicon using the plasmidic DNA preparation strongly suggested that it was plasmid-borne. In addition, they revealed that the *dsrF* and *dsrR* primers pair designed by us are useful for the detection of *dsr* genes in a similar way to those degenerate oligonucleotides previously used for the detection of genes encoding glucosyltransferases of *Lactobacilli* (Kralj et al., 2003).

Analysis of a plasmidic DNA preparation from *Lb. sakei* MN1 in agarose gel revealed the presence of two groups of two bands each (Figure 1A). In addition, it was detected that exposure of the DNA preparations to repeating freezing (at  $-80^{\circ}\text{C}$ ) and thawing (at  $4^{\circ}\text{C}$ ) cycles resulted in alteration of the intensity of the bands: an increase in the two upper bands accompanied by a decrease in the two faster migrating ones (results not shown). The sizes of the bands were inferred from their migration using a calibration curve (Figure 1B) generated with the plasmids of the *E. coli*



V517 strain (**Figure 1A**, lane S), and the overall results indicated that bands with higher mobility could be the covalently closed circles of two plasmids of ~12 and 14 kbp named respectively pMN1 and pMN2, whereas bands with less mobility could be the

open circle forms of those plasmids. To disclose the location of *dsrLS*, the 695-bp amplicon was used as a probe for Southern blot hybridization and the blot revealed the presence of two hybridization bands corresponding to the two proposed pMN1 forms (**Figure 1A**).

## Characterization of Plasmid pMN1

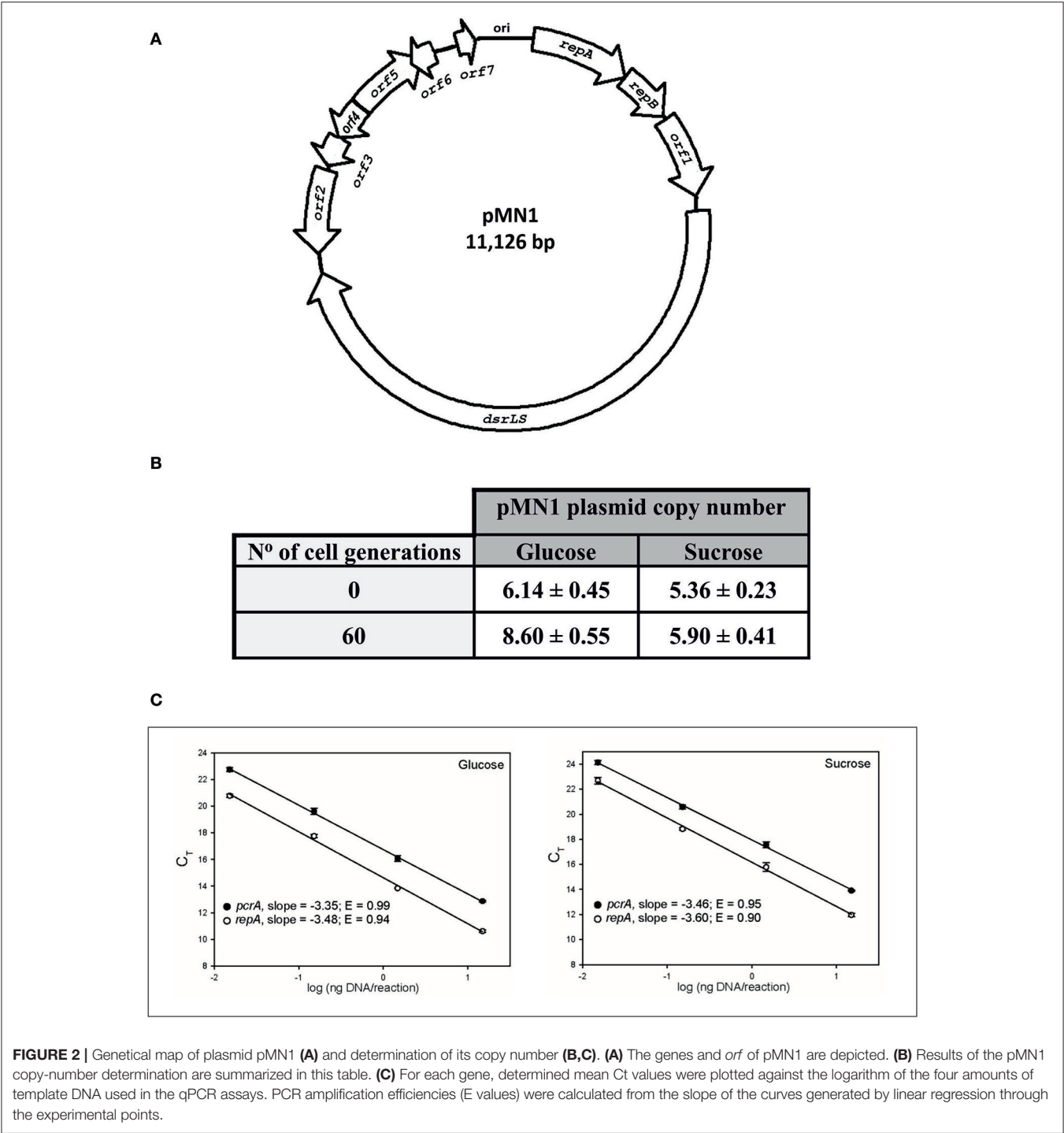
Following DNA sequencing, the size of pMN1 plasmid was estimated as 11,126 kbp. A genetic map of pMN1 is depicted in **Figure 2A**. Blast alignment of its DNA sequence with those deposited in Genbank lead to confirm the existence of the 5,304-bp *dsrLS* gene. Moreover, in pMN1, a 1,674-bp replicon homologous to the replicons of the pUCL287 plasmid family was identified that replicates bidirectionally by theta mechanism (Benachour et al., 1995). This region contains the origin of replication and the *repA* and *repB* genes. In addition, in pMN1, seven other open reading frames were identified and designated ORFs 1-7, which could encode hypothetical proteins. Blast analysis of inferred amino acid sequence of the ORFs with those deposited in the Non-redundant protein sequences database revealed that *orf1*, *orf3*, *orf4*, and *orf5* could encode respectively a truncated type I restriction endonuclease subunit R, a transcriptional regulator belonging to the XRE family, a RelE type toxin addition module and a site-specific integrase.

Furthermore, the copy number of the plasmid was investigated by real time-qPCR by analysis of the *repA* gene vs. that of the chromosomal housekeeping and monocopy *pcrA* gene. *Lb. sakei* MN1 cultures were maintained in exponential growth phase in either MRSG or MRSS and DNA preparations of cultures grown to exponential phase once (from the glycerol stock) or for 60 successive generations (by six subsequent 1/1,000 dilutions) were analyzed. The results are summarized in **Figure 2B**, and revealed that pMN1 is a low-copy-number plasmid ( $\sim 6.5 \pm 1.5$  copies per genome equivalent), which maintains its copy number over at least 60 generations, and that its copy number is not significantly affected by conditions required for dextran synthesis.

## Gene Expression of *dsrLS*

We have previously detected that *Lb. sakei* MN1 is unable to synthesize the dextran in the absence of sucrose (Nácher-Vázquez et al., 2017). This could be due not only to the lack of the substrate for the polymer synthesis, but also to the fact that *dsrLS* gene expression requires an induction mediated by the disaccharide as it has been detected in other LAB (Neubauer et al., 2003). In addition, inspection of the pMN1 DNA sequence indicated that transcription of *dsrLS* could be driven from more than one promoter located upstream of this gene. Thus, total RNA preparations were obtained from *Lb. sakei* MN1 cultures grown in medium containing either glucose or sucrose and five RT-PCR reactions were performed to generate the amplicons showed in **Figure 3A**. These amplicons contain regions of more than one gene and their corresponding intergenic regions (amplicons 1, 2, 3, and 5) or only a region of *dsrLS* (amplicon 4). The results revealed that four reactions generated the expected sizes 1, 2, 3, and 4 amplicons, which included regions located upstream of, or within, the *dsrLS* gene (**Figure 3B**). In addition, the



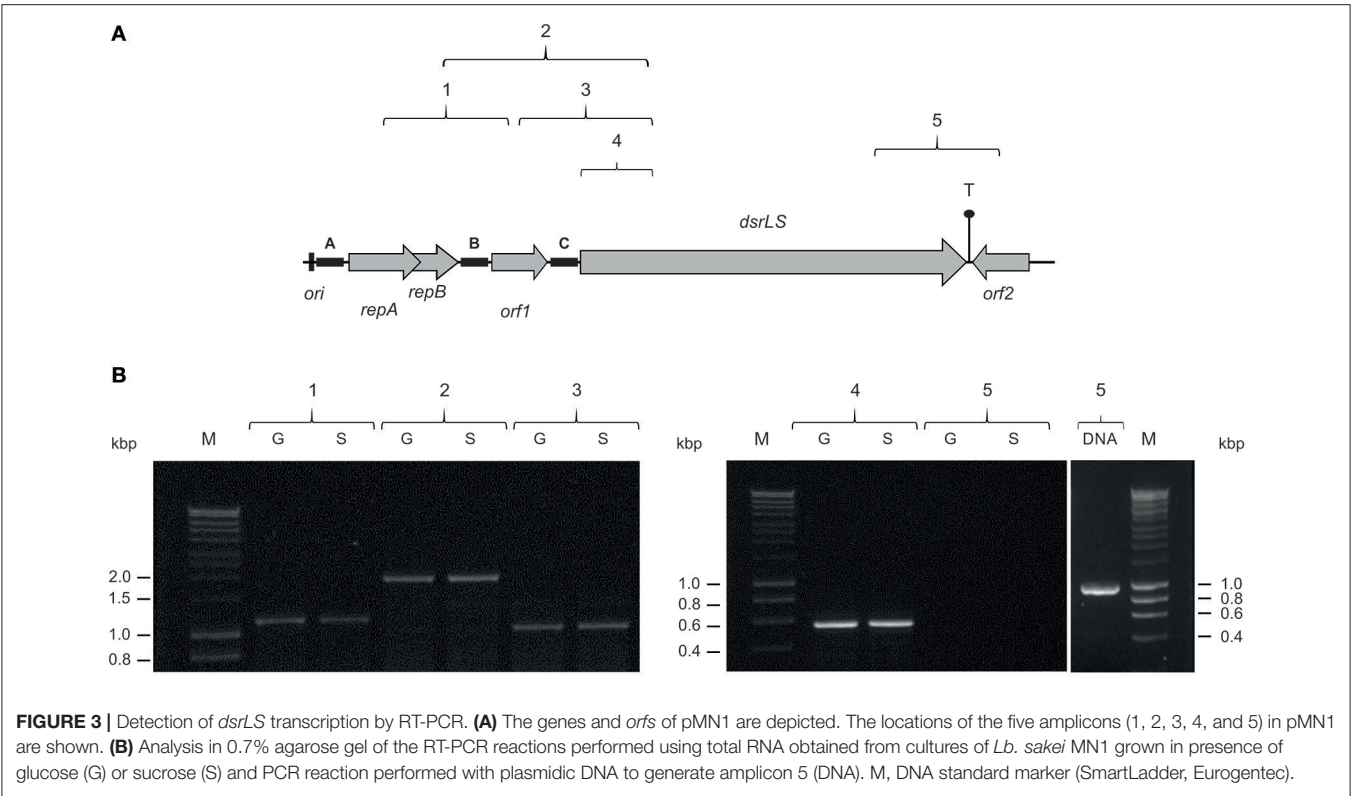


quantification of the intensity of the amplicons (results not shown) indicated that the mRNA levels were very similar in cultures grown in CDMG or CDMS media. The fifth reaction did not reveal the amplicon 5 (**Figure 3B**), which carries the 3'-end of *dsrLS* and downstream regions (**Figure 3A**). This latter result was not due to the primers used, since the expected amplicon was obtained using the plasmidic DNA preparation as substrate (**Figure 3B**). In addition, the negative result was expected due to

the convergent polarity of *dsrLS* and *orf2* and to the existence of a putative bidirectional transcriptional terminator located between the 3'-ends of the two genes, which predicted the lack of their co-transcription.

**Identification of Promoter Regions in pMN1**

As expected, the RT-PCR analysis revealed that *repA* and *repB* were co-transcribed, but in addition showed the existence of



**FIGURE 3 |** Detection of *dsrLS* transcription by RT-PCR. **(A)** The genes and *orfs* of pMN1 are depicted. The locations of the five amplicons (1, 2, 3, 4, and 5) in pMN1 are shown. **(B)** Analysis in 0.7% agarose gel of the RT-PCR reactions performed using total RNA obtained from cultures of *Lb. sakei* MN1 grown in presence of glucose (G) or sucrose (S) and PCR reaction performed with plasmidic DNA to generate amplicon 5 (DNA). M, DNA standard marker (SmartLadder, Eurogentec).

transcripts including *dsrLS* and upstream genes. Therefore, detection of DNA regions involved in *dsrLS* expression was approached. We have previously developed the pRCR promoter probe plasmid (Figure S1) based on the pSH71 replicon, which replicates *via* a rolling circle mechanism and carries a mCherry-coding gene (*mrfp*) optimized for expression of the fluorescent protein in LAB. Moreover, we had shown functionality of this replicon concomitant with successful expression of this *mrfp* in *L. lactis* (Mohedano et al., 2015) and *Lb. sakei* MN1 (Nácher-Vázquez et al., 2017). Thus, to detect promoter regions that could drive transcription of *dsrLS* in *L. lactis* and *Lb. sakei*, A, B, and C DNA fragments, carrying intergenic regions located upstream of the *repA*, *orf1*, and *dsrLS* genes, respectively (Figure 3), were cloned upstream of the *mrfp* gene into the pRCR plasmid, generating the recombinant plasmids pRCR13, pRCR14, and pRCR15. Thus, these plasmids carry putative transcriptional fusions to the mCherry-coding gene (Figure S1).

The clonings were performed in *L. lactis* MG1363 and then the plasmids were transferred to *Lb. sakei* MN1. Afterwards, expression of the mCherry in the two hosts was monitored by fluorescent spectroscopy and microscopy. *L. lactis* cannot grow in media containing sucrose as the only carbon source. Therefore, expression of mCherry was monitored during growth in M17G and M17GS. The results obtained during the early and late stationary phases are shown in Table 3 and Figure S2. As expected, MG1363 did not show fluorescence. Moreover, fluorescence was observed only in cultures of MG1363[pRCR13] and MG1363[pRCR15] and the promoter regions present in

**TABLE 3 |** Fluorescent detection of promoter regions in *L. lactis* by translational fusions to the *mrfp* gene.

Growth phase  Growth medium  Bacteria	Initial stationary phase		Late stationary phase	
	M17G	M17GS	M17G	M17GS
<i>L. lactis</i> MG1363	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.1	0.10 ± 0.0
<i>L. lactis</i> MG1363 [pRCR13]	0.7 ± 0.5	0.8 ± 0.1	16.6 ± 0.8	12.5 ± 1.2
<i>L. lactis</i> MG1363 [pRCR14]	0.2 ± 0.1	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
<i>L. lactis</i> MG1363 [pRCR15]	0.6 ± 0.2	0.9 ± 0.1	10.2 ± 0.8	10.4 ± 1.6

<sup>a</sup>The specific fluorescence is depicted and it was calculated as the ratio of the detected fluorescence (10×) and the bacterial biomass estimated from the A<sub>600</sub> of the culture.

them were designated P1 and P2, respectively. Nevertheless, the levels of fluorescence in these strains were low and only significantly detected at late stationary phase. This pattern of expression had been previously observed when the *mrfp* gene was expressed in *L. lactis* MG1363 under control of a lactococcal promoter (Garcia-Cayuela et al., 2012), and it could indicate that the mCherry protein requires a long period of maturation prior to emit fluorescence in this host.

In the case of *Lb. sakei*, strains were grown in MRSG and MRSS and the presence of pRCR13 and pRCR15, and not of pRCR14, conferred fluorescence to *Lb. sakei* MN1, in both exponential and stationary phases (Table 4 and Figure 4). The levels of specific fluorescence indicated that P1 is stronger than P2

**TABLE 4 |** Fluorescent detection of promoter regions in *Lb. sakei* by expression of translational fusions to the *mrfp* gene.

Growth phase	Growth medium	<sup>a</sup> Middle exponential phase		<sup>a</sup> Late stationary phase	
		Glucose	Sucrose	Glucose	Sucrose
<i>Lb. sakei</i> MN1		1.3 ± 0.6	0.9 ± 0.1	0.5 ± 0.1	0.4 ± 0.5
<i>Lb. sakei</i> MN1[pRCR12]		355 ± 30	351 ± 18	227 ± 14	139 ± 5.5
<i>Lb. sakei</i> MN1[pRCR13]		142 ± 14	76 ± 7.5	149 ± 11	68 ± 2.2
<i>Lb. sakei</i> MN1[pRCR14]		4.6 ± 0.4	2.0 ± 0.2	5.5 ± 0.3	1.6 ± 0.1
<i>Lb. sakei</i> MN1[pRCR15]		73 ± 2.6	28 ± 2.0	40 ± 0.9	8.7 ± 0.4

<sup>a</sup>The specific fluorescence is depicted and it was calculated as the ratio of the detected fluorescence (10×) and the bacterial biomass estimated from the A<sub>600</sub> of the culture.

and that they are weaker than the pneumococcal Px promoter, which drives transcription of *mrfp* in *Lb. sakei* MN1[pRCR12] (Table 4) (Nácher-Vázquez et al., 2017).

In addition, at the exponential phase, the specific fluorescence in MN1[pRCR13] and MN1[pRCR15] strains was two-fold-higher in MRSG than in MRSS. These results confirmed that the sucrose present in the medium is not an inducer of the *dsrLS* gene.

## DISCUSSION

The enzymes responsible for the HoPS synthesis are glycosyl hydrolases, extracellular polymerases that utilize the energy of the glucosidic bond of sucrose to link molecules of glucose. If they synthesize α-D-glucans are called glucansucrases and according to the CAZy classification (<http://www.cazy.org>), are members of the GH70 family. Among them, dextranucrases synthesize dextran and here we have characterized the *Lb. sakei* MN1 *dsrLS* gene of 5,304-bp, which encodes the DsrLS composed of 1,767 amino acids (aa). Analysis of these sequences with the BLAST program vs. those deposited in the NCBI databases revealed homologies with other bacterial genes and with their gene products. The highest homology was detected with the 5,094-bp *gtf1624* gene from *Lb. curvatus* TMW1624 and its product, the dextranucrase GTF1624 of 1,697 aa (Rühmkorf et al., 2013). In addition, the *dsrLS* gene and DsrLS also exhibited high homology with the *gtfkg15* gene of *Lb. sakei* Kg15 of 4,788-bp and its product GTFKg15 of 1,595 aa (Kralj et al., 2004a). An alignment of the three proteins is presented in Figure S3.

It has been demonstrated that the EPS synthesized by GTF1624 (Rühmkorf et al., 2013) and GTFKg15 (Kralj et al., 2004a) are α-(1-6)-glucans with a low percentage of substitutions at positions O-3, like the EPS synthesized by *Lb. sakei* MN1 (Nácher-Vázquez et al., 2015). This fact strongly supports that indeed DsrLS is the enzyme responsible for the synthesis of the MN1 dextran.

In all glucansucrases, including dextranucrases, exist: (i) a N-terminal, variable region, (ii) the catalytic domain, and (iii) a C-terminal, so-called “glucan binding” domain (van Hijum et al., 2006). All these regions were identified in DsrLS and

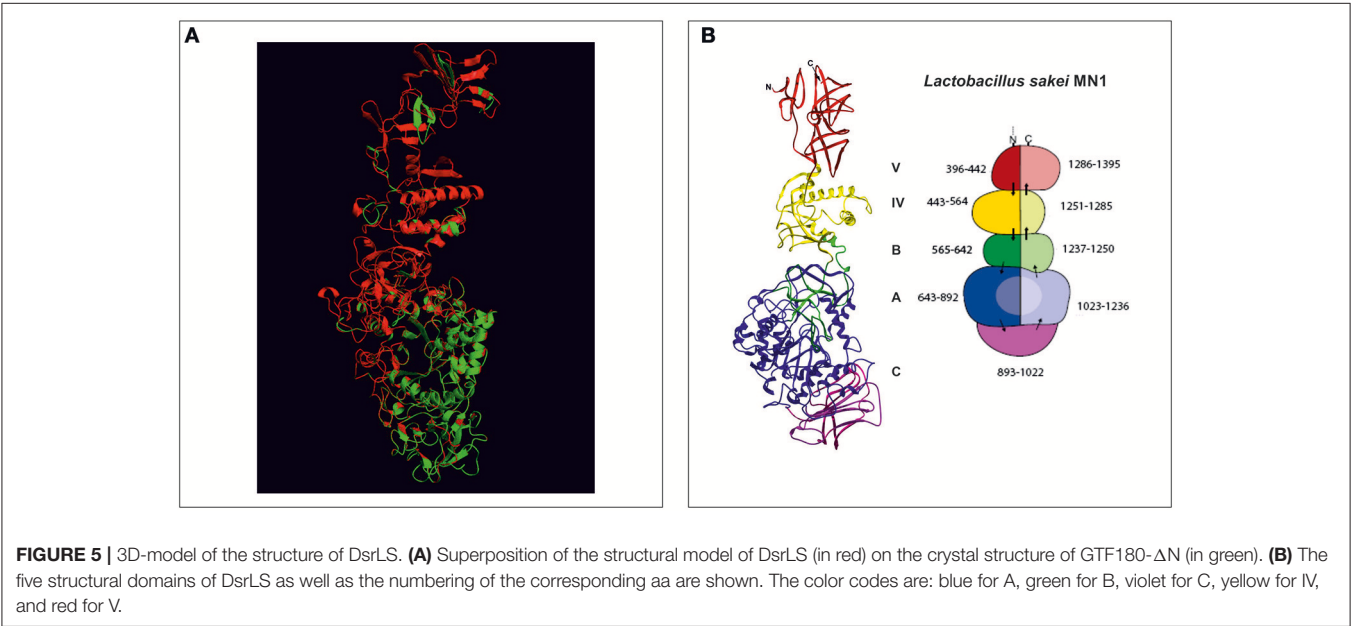
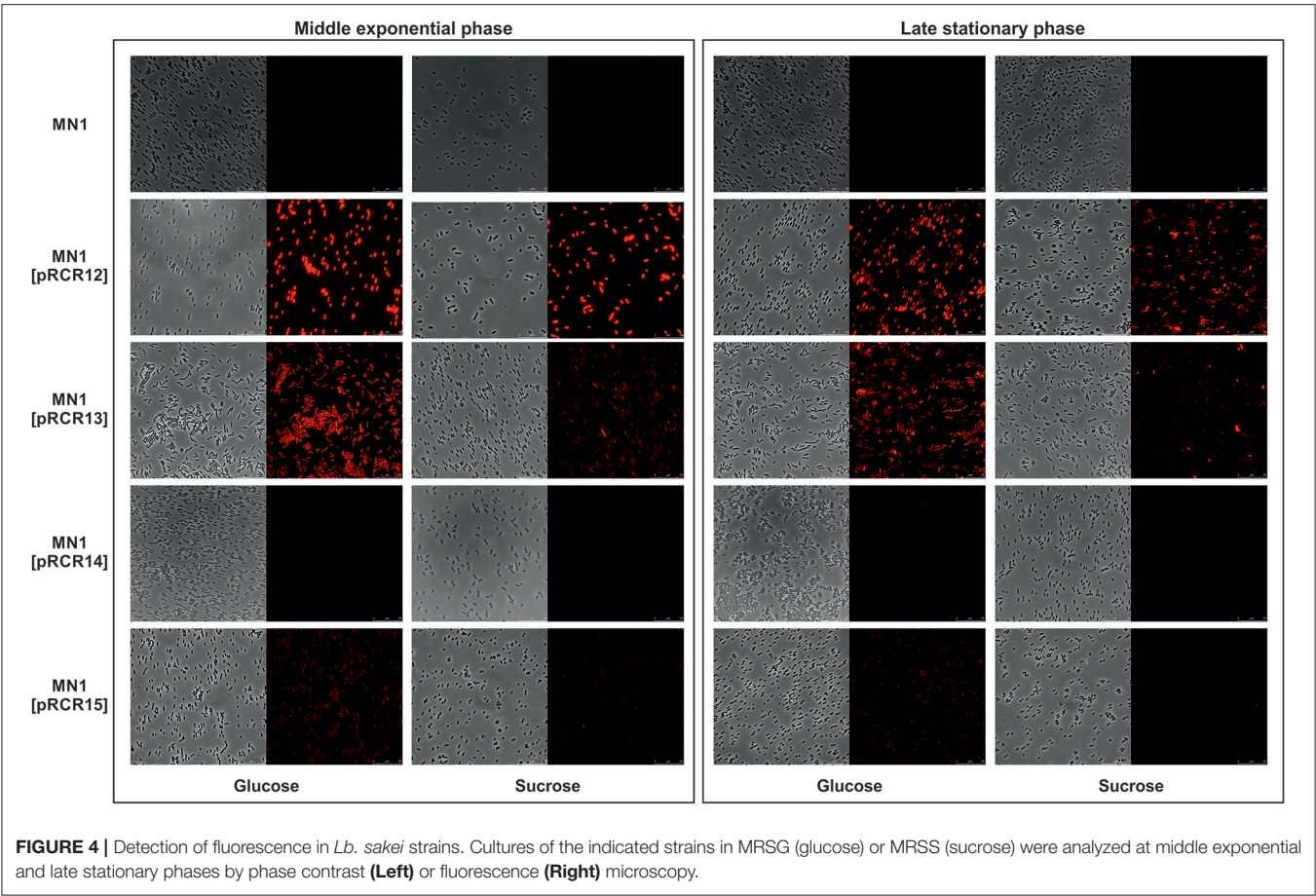
include the following aa: (i) 49–390, (ii) 391–1,154, and (iii) 1,155–1,767. The glucansucrases are extracellular enzymes, and at the N-terminus of DsrLS it was found a sequence characteristic of the leader peptides of Gram-positive bacteria (1–48 aa), also present in GTF1624 and GTFKg15. In addition, the difference in the number of aa of the three dextranucrases is due in one hand to the absence in GTF1624 of 70 aa present in the C-terminal region of DsrLS (residues 1,541–1,610). On the other hand, DsrLS has 172 aa more than its homolog of *Lb. sakei* Kg15, being 145 of them located at the C-terminal region (1,485–1,629) and 27 aa at its N-terminal variable region (residues 64–89 and 813). Consequently, the greatest divergence of DsrLS with both GTF1624 and GTFKg15 is located at its C-terminal domain.

Currently, the 3D-structure of any entire glucansucrase has not been determined, but partial structures of: (i) the DSR-E-ΔN of *Lc. mesenteroides* NRRLB-1299 (PDB 3TTQ), (ii) the GTF-S1 of *Streptococcus mutants* (PDB 3AIE), (iii) the GTFA-ΔN of *Lb. reuteri* 121 (PDB 4AMC), and (iv) the GTF180-ΔN of *Lb. reuteri* 180 (PDB 3KLK, 4AYG, and 3HZ3) have been solved by X-ray diffraction analysis of crystals. From the crystal structure of these proteins, it has been established that there are five structural domains designated A, B, C, IV, and V (Leemhuis et al., 2013). The A, B, and C domains have been named following the nomenclature of the structurally homologous domains of the GH13 family of α-amylases. The domains IV and V have not homologs in GH13 and for this reason have been named with a different nomenclature (Vujicic-Zagar et al., 2010). These domains are not adjacent in the primary structure of the proteins, and are located with a “U” distribution and a pattern V, IV, B, A, C, A, B, IV, and V.

The aa sequence of DsrLS (residues 396–1,395) has an identity of 51% with that of the GTF180-ΔN of *Lb. reuteri* 180 (PDB 3HZ3) (Figure S4). Thus, it was possible to develop a model of the 3D-structure of DsrLS lacking the N- and the C-terminal regions (Figure 5), which predicts that the *Lb. sakei* enzyme has the same domains as its homonymous of *Lb. reuteri* (Figure 5B). The A domain is a barrel (β/α)<sub>8</sub> and contains the catalytic site of the enzyme, including an amino acid triad composed of two aspartate and one glutamate residues, which are involved in the formation of a covalent glucosyl-enzyme intermediate, the key step in the transfer of D-glucosyl units. From this intermediate, the glucosyl unit is transferred to the acceptor (the growing dextran molecule) by a processive catalytic mechanism. Thus, superposition of the 3D-model of DsrLS on the co-crystal of GTF180-ΔN and sucrose indicates that D678, D789, and E716 constitute the catalytic triad of the *Lb. sakei* MN1 dextranucrase (Figure 6).

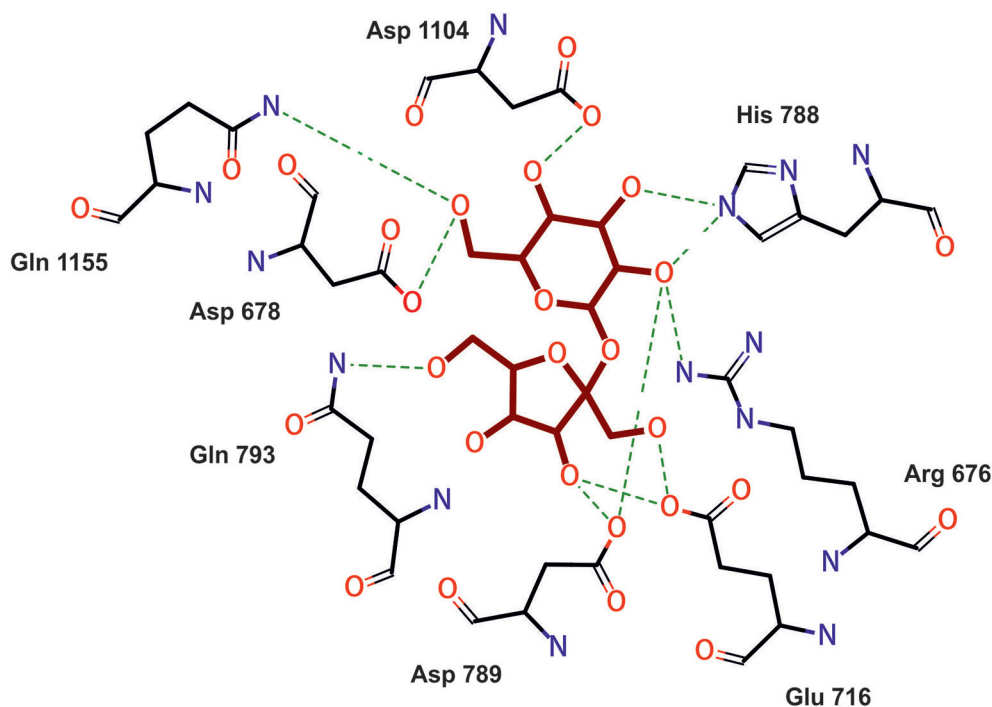
It has been shown that calcium is essential for GTFA-ΔN (Kralj et al., 2004b) and GTF180-ΔN (Vujicic-Zagar et al., 2010) activities. The B domain located adjacent to the A domain seems to be essential for glucansucrases activity because (i) calcium binding site includes aa from A and B domains, and (ii) some elements of the B domain contribute to stabilize the conformation of the A domain. Concerning to the C domain, although is conserved in all glucansucrases, its function is still unknown. The IV domain connects the B and V domains and seems to act as a hinge to bring the V domain close to the catalytic domain (Ito et al., 2011). The V domain is constituted





by the N- and C-terminal regions, which include a series of structural modules with two or three  $\beta 2/\beta 3$  units containing  $\sim 20$  aa and arranged in a regularly repeating fashion, resulting in a  $\beta$ -solenoid fold. In some glucansucrases these modules include YG repeats (containing a tyrosine/glycine motif) (Leemhuis et al., 2013) and DsrLS carries YG in both terminal regions.





**FIGURE 6 |** Model of amino acids interaction with sucrose (magenta) at the catalytic center of DsrL. Model obtained by superimposing the model structure of DsrLS on the co-crystal of GTF180- $\Delta$ N and sucrose.

In *Lactobacillus*, the variable N-terminal region of glucansucrases contains 200–700 aa and mutations or deletions of this region can alter the functions of the proteins. Thus, in GTFA, deletion of this region affects the interplay between the hydrolytic and transglycosidase activities (Kralj et al., 2004b). The C-terminal domain contains  $\sim$ 300 aa and the function of this region is still unknown. In GTFA, its deletion diminishes affinity for sucrose (Kralj et al., 2004a). The repetitions at the C-terminal region of glucansucrases have homology with motives for binding to bacterial cell wall present in choline-binding proteins, toxins, and other bacterial surface proteins (Leemhuis et al., 2013). However, the function of the C-terminal region is currently unknown, although its implication in several functions has been proposed: (i) polymerization or glucan structure, (ii) transfer of products to the catalytic center, and (iii) anchoring of the protein to the bacterial surface.

The *Lb. sakei* DsrLS and GTFKg15 enzymes only differ in the number of aa at their C-terminal (307 and 162 residues, respectively) and N-terminal (342 and 315 residues, respectively) regions, and the *Lb. reuteri* GTF180 lacks the C-terminal region.

These differences could be related to the enzymes processivity, since the dextrans synthesized in fermentation conditions by *Lb. sakei* MN1, *Lb. sakei* Kg15 and *Lb. reuteri* 180 have molecular masses of  $1.7 \times 10^8$  Da (Zarour et al., 2017),  $2.7 \times 10^7$  Da, and  $3.6 \times 10^6$  (Kralj et al., 2004a), respectively.

Here, we have demonstrated that the gene encoding DsrLS is carried by the 11,126 kbp pMN1 plasmid. Plasmidic localization of the genetic determinants for the production of EPS has

been described previously (Wang and Lee, 1997). It has been determined, by plasmid curing, that the production of a HePS in *Lactobacillus casei* CG11 depends on the presence of a plasmid of 30 kbp (Kojic et al., 1992). Also, the *gtf* gene encoding the GTF glycosyltransferase, which synthesizes a O2-substituted (1,6)- $\beta$ -D-glucan, has been identified in plasmids of *Pediococcus* and *Lactobacilli* strains (Werning et al., 2006). Concerning to dextran synthesis, it was shown that production of the polymer by two *Lactobacillus* strains isolated from meat was impaired upon curing of a 11 kbp plasmid (Ahrné et al., 1989), which could be identical or similar to pMN1. However, as far as we know, this is the first time that a plasmid carrying a gene encoding a dextranucrase has been completely sequenced.

Homology of pMN1 with other plasmids revealed that it belongs to a plasmid family whose prototype is pUCL287, which replicates *via* theta-mode. The pMN1 replicon includes two genes, *repA* and *repB*, which should encode the RepA and RepB proteins, involved in the initiation of plasmid replication and regulation of plasmid copy number. Thus, RepB could be responsible for the segregational stability of the plasmid and in fact the results obtained here indicate that low copy number pMN1 is stably inherited. In addition, the putative RelE toxin encoded by *orf4*, whose expression is probably regulated by the product of *orf3*, could be other mechanism to eliminate the bacterial population that has lost the plasmid.

We have previously shown that *Lb. sakei* MN1 utilizes very efficiently sucrose with production of dextran and without accumulation of glucose, not affecting the growth rate and

resulting in a higher biomass than when the growth medium was supplemented with glucose instead of sucrose (Nácher-Vázquez et al., 2017). Thus, it is not strange that the plasmid is segregationally stable because, in addition to encode RepB, it is not a burden for the cells and rather the bacteria is benefited by its presence. Upstream of *repA*, several iterons were identified: (i) 4 direct repeats of 11-bp (Figure 7A), which according to the studies performed with pUCL287 (Benachour et al., 1997) constitute the replication origin of the plasmid and presumably the RepA binding site, and (ii) 5 direct repeats of 22-bp, which could be involved in partitioning or incompatibility processes.

The BLAST analysis of the pMN1 DNA sequences vs. those deposited in the NCBI databases revealed homologies of ORF 2, 3, 4, 5, 6, and 7 with ORF 13, 12, 11, 10, 9, and 8 of the plasmid pRV500 from *Lb. sakei* RV332 (Alpert et al., 2003), which belongs to the same plasmid family as pMN1. The major difference between pRV500 and pMN1 is that the first one carries the genes encoding a restriction-modification type I system instead of the pMN1 *dsrLS* gene and its preceding *orf1*. This last one appears to be a truncated sequence of a gene that initially encoded a specific deoxyribonuclease (R protein), belonging to a type I restriction and modification system different to that of pRV500 (Figure 7B).

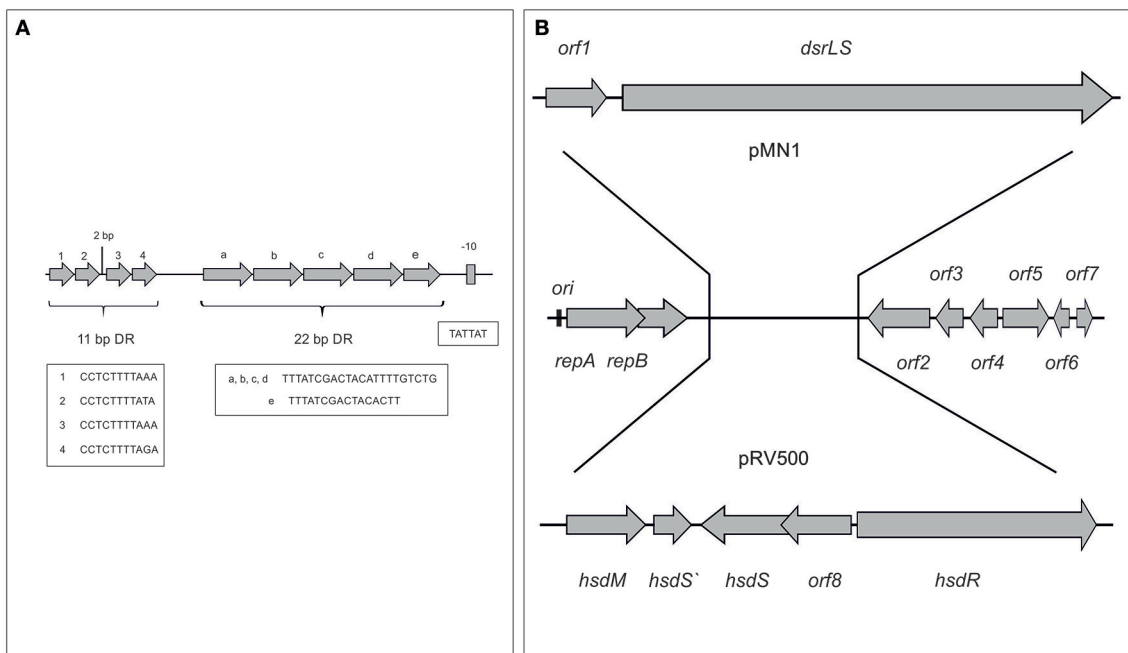
Most lactobacilli carry more than one plasmid and both pMN1 and pRV500 have been detected in *Lb. sakei* strains isolated from meat products, although in different countries. Thus, these plasmids seem to be derived from a parental plasmid composed of a replicon and the *orfs* 2–7, which subsequently by means of transposition processes incorporated modules that allow the synthesis of the dextran or a system of restriction and modification. Possibly, the acquisition of one or other module

and its fixation is due to the selective advantage that supposes for their hosts against the environmental stress or the infection by bacteriophages. Moreover, the product of the *orf5*, a site-specific integrase could be involved in the process of modules exchange.

Depending on the requirement for sucrose utilization, expression of the dextransucrases could be constitutive or inducible. To date, it has been determined that its synthesis is constitutive in *Streptococcus* (Janda and Kuramitsu, 1978; Wenham et al., 1979), while it is inducible in presence of sucrose in *Leuconostoc* (Neely and Nott, 1962; Funane et al., 1995), although the molecular mechanisms of this induction are unknown. In *Weissella*, levels of dextransucrase detected in cultures of *Weissella cibaria* and *Weissella confusa* grown in presence of different sugars indicate constitutive expression of their Dsr enzymes (Bounaïx et al., 2010). Determination of levels and activity of glucansucrases in *Lactobacillus* also points to a constitutive expression. This is the case of the glucan-producing *Lb. reuteri* TMW1106 (Schwab et al., 2007), the reuteran-producing *Lb. reuteri* 121 (Kralj et al., 2004a), and the dextran-producing *Lb. reuteri* 180 and *Lactobacillus parabuchneri* 33 (Kralj et al., 2004a).

In this work we have determined by RT-PCR analysis that the expression of the *dsrLS* gene of *Lb. sakei* MN1 did not increase when sucrose is present in the culture medium. Thus, the disaccharide is not an inducing agent of the *dsrLS* expression.

The transcriptional fusions generated in this work revealed two promoter regions designated P1 and P2, which drive expression of *dsrLS*. Inspection of the pMN1 sequences cloned in pRCR13 revealed a DNA sequence TATtAT (Figure 7A) which only deviates a nucleotide from the canonical –10 promoter



**FIGURE 7 |** Origin of replication of pMN1 (A) and model of generation of pMN1 and pRV500 from a putative parental plasmid (B). In (A) the sequences of the iterons and of the putative promoter of the *repA-repB-ori1-dsrLS* operon are depicted. In (B) the genes and the *orfs* of plasmids are depicted.

sequence. This sequence is located 16 nucleotides upstream of *repA*, and could be P1. Also, within the pMN1 insert of pRCR15, a –10 extended promoter region (TGTTATtAT) with only one mismatch was observed 82 nucleotides upstream of *dsrLS*, that could correspond to P2. No –35 promoter region was detected for either P1 or P2. In addition, in late stationary phase cultures of *Lb. sakei* and not of *L. lactis* carrying pRCR15 a four-fold higher fluorescence levels was detected, when grown in medium containing only glucose (Tables 3, 4). Taking in consideration that *Lb. sakei* MN1 is the native background for the P2 promoter, it is feasible that a negative effector of the P2 promoter could be present or encoded by pMN1, pMN2 or the chromosome. Moreover, lower fluorescence was detected in *Lb. sakei* exponential cultures expressing mCherry from P1 or P2, when they were grown in MRSS. This fact could be due to the presence of dextran in cultures grown in MRSS, which could cause a shielding that masks fluorescence. However, analysis of individual cells by fluorescence microscopy also revealed a lower fluorescence in bacteria grown in MRSS medium (Figure 4), despite cells were washed repeatedly to remove the dextran prior to the analysis. In addition, difference in fluorescence levels between bacteria grown in MRSG or MRSS was not observed during exponential phase in MN1 strain when carried pRCR12 plasmid. Therefore, an alternative hypothesis is that the high synthesis of dextran (around 10 g L<sup>-1</sup> in MRSS), although it does not seem to affect bacterial growth (Nácher-Vázquez et al., 2017), could finally be an energetic burden for the bacterium and the dextran itself could induce the activation of an inhibitory mechanism of its own synthesis at the transcriptional level. Another hypothesis, more feasible, is that the fructose synthesized as a consequence of the hydrolysis of sucrose, is the effector of inhibition. However, further studies will be necessary in order to elucidate the real cause of the detected effect.

An unexpected co-transcription of *dsrLS* with the *repA* and *repB* genes has been detected here. We believe that this is the first described case of co-transcription of genes involved in plasmid replication and a gene encoding a protein involved in the synthesis of EPS. However, this is not the first case in the literature of co-transcription of replication genes together with other genes, since the plasmidic *dysI* gene (encoding the immunity factor of the streptococcal bacteriocin dysgalactin) is transcribed as part

of the *copG-repB-dysI* replication-associated operon (Swe et al., 2010).

Finally, we would like to highlight the multicopy state of *dsrLS* in the pMN1 stable plasmid, its expression from two promoters not induced by sucrose, and the apparently processive DsrLS, which synthesizes a high-molecular mass dextran with antiviral and immunomodulatory activities (Nácher-Vázquez et al., 2017), as well as rheological properties (Zarour et al., 2017). Thus, all these facts support the potential of *Lb. sakei* MN1 and its dextran for multiple industrial applications including those in functional food.

## AUTHOR CONTRIBUTIONS

MN-V contributed to all parts of the experimental work and wrote a draft of the manuscript. JR-M performed the plasmid characterization. MM contributed to the transcriptional gene expression analysis. GdS performed the bioinformatics analysis of pMN1 plasmid, interpreted this analysis, and revised the manuscript. RA participated in study conception and corrected the manuscript. PL participated in study conception, data interpretation, and generated the final version of the manuscript. All authors have read and approved the final manuscript.

## FUNDING

This work was supported by the Spanish Ministry of Economy and Competitiveness (grant AGL2015-65010-C3-1-R).

## ACKNOWLEDGMENTS

We thank Dr. Stephen W. Elson for critical reading of the manuscript. We thank Dr. Mario García Lacoba for his valuable assistance in the modeling of DsrLS.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02281/full#supplementary-material>

## REFERENCES

- Ahrné, S., Molin, G., and Ståhl, S. (1989). Plasmids in *Lactobacillus* strains isolated from meat and meat products. *Syst. Appl. Microbiol.* 11, 320–325. doi: 10.1016/S0723-2020(89)80031-1
- Alpert, C. A., Crutz-Le Coq, A. M., Malleret, C., and Zagorec, M. (2003). Characterization of a theta-type plasmid from *Lactobacillus sakei*: a potential basis for low-copy-number vectors in lactobacilli. *Appl. Environ. Microbiol.* 69, 5574–5584. doi: 10.1128/AEM.69.9.5574-5584.2003
- Aman, A., Siddiqui, N. N., and Qader, S. A. U. (2012). Characterization and potential applications of high molecular weight dextran produced by *Leuconostoc mesenteroides* AA1. *Carbohydr. Polym.* 87, 910–915. doi: 10.1016/j.carbpol.2011.08.094
- Amari, M., Arango, L. F., Gabriel, V., Robert, H., Morel, S., Moulis, C., et al. (2013). Characterization of a novel dextranucrase from *Weissella confusa* isolated from sourdough. *Appl. Microbiol. Biotechnol.* 97, 5413–5422. doi: 10.1007/s00253-012-4447-8
- Anastasio, M., Pepe, O., Cirillo, T., Palomba, S., Blaiotta, G., and Villani, F. (2010). Selection and use of phytate-degrading LAB to improve cereal-based products by mineral solubilization during dough fermentation. *J. Food Sci.* 75, M28–M35. doi: 10.1111/j.1750-3841.2009.01402.x
- Badel, S., Bernardi, T., and Michaud, P. (2011). New perspectives for lactobacilli exopolysaccharides. *Biotechnol. Adv.* 29, 54–66. doi: 10.1016/j.biotechadv.2010.08.011
- Benachour, A., Frere, J., Boutibonnes, P., and Auffray, Y. (1995). Characterization and replication mode determination of the minimal replicon of *Tetragenococcus halophila* ATCC33315 plasmid pUCL287. *Biochimie* 77, 868–874. doi: 10.1016/0300-9084(95)90005-5
- Benachour, A., Frere, J., Flahaut, S., Novel, G., and Auffray, Y. (1997). Molecular analysis of the replication region of the theta-replicating plasmid pUCL287

- from *Tetragenococcus (Pediococcus) halophilus* ATCC33315. *Mol. Gen. Genet.* 255, 504–513. doi: 10.1007/s004380050523
- Berthier, F., Zagorec, M., Champomier-Vergès, M., Ehrlich, S. D., and Morel-Deville, F. (1996). Efficient transformation of *Lactobacillus sake* by electroporation. *Microbiology* 142, 1273–1279. doi: 10.1099/13500872-142-5-1273
- Bounaix, M. S., Robert, H., Gabriel, V., Morel, S., Remaud-Simeon, M., Gabriel, B., et al. (2010). Characterization of dextran-producing *Weissella* strains isolated from sourdoughs and evidence of constitutive dextranucrase expression. *FEMS Microbiol. Lett.* 311, 18–26. doi: 10.1111/j.1574-6968.2010.02067.x
- Capozzi, V., Russo, P., Dueñas, M. T., López, P., and Spano, G. (2012). Lactic acid bacteria producing B-group vitamins: a great potential for functional cereals products. *Appl. Microbiol. Biotechnol.* 96, 1383–1394. doi: 10.1007/s00253-012-4440-2
- Chenoll, E., Macián, M. C., Elizaquível, P., and Aznar, R. (2007). Lactic acid bacteria associated with vacuum-packed cooked meat product spoilage: population analysis by rDNA-based methods. *J. Appl. Microbiol.* 102, 498–508. doi: 10.1111/j.1365-2672.2006.03081.x
- de Man, J. C., Rogosa, M., and Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J. Appl. Microbiol.* 23, 130–135. doi: 10.1111/j.1365-2672.1960.tb0188.x
- Dimopoulou, M., Bardeau, T., Ramonet, P. Y., Miot-Certier, C., Claisse, O., Doco, T., et al. (2015). Exopolysaccharides produced by *Oenococcus oeni*: from genomic and phenotypic analysis to technological valorization. *Food Microbiol.* 53, 10–17. doi: 10.1016/j.fm.2015.07.011
- Dornan, S., and Collins, M. (1987). High efficiency electroporation of *Lactococcus lactis* subsp. *lactis* LM0230. *Lett. Appl. Microbiol.* 11, 62–64. doi: 10.1111/j.1472-765X.1990.tb01275.x
- Funane, K., Yamada, M., Shiraiwa, M., Takahara, H., Yamamoto, N., Ichishima, E., et al. (1995). Aggregated form of dextranucleases from *Leuconostoc mesenteroides* NRRL B-512F and its constitutive mutant. *Biosci. Biotechnol. Biochem.* 59, 776–780. doi: 10.1271/bbb.59.776
- García-Cayuela, T., de Cadinanos, L. P., Mohedano, M. L., de Palencia, P. F., Boden, D., Wells, J., et al. (2012). Fluorescent protein vectors for promoter analysis in lactic acid bacteria and *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 96, 171–181. doi: 10.1007/s00253-012-4087-z
- Gasson, M. J. (1983). Plasmid complements of *Streptococcus lactis* NCDO 712 and other streptococci after protoplast-induced curing. *J. Bacteriol.* 154, 1–9.
- Ito, K., Ito, S., Shimamura, T., Weyand, S., Kawarasaki, Y., Misaka, T., et al. (2011). Crystal structure of glucanucrase from the dental caries pathogen *Streptococcus mutans*. *J. Mol. Biol.* 408, 177–186. doi: 10.1016/j.jmb.2011.02.028
- Janda, W. M., and Kuramitsu, H. K. (1978). Production of extracellular and cell-associated glucosyltransferase activity by *Streptococcus mutans* during growth on various carbon sources. *Infect. Immun.* 19, 116–122.
- Kojic, M., Vujcic, M., Banina, A., Cocconcelli, P., Cerning, J., and Topisirovic, L. (1992). Analysis of exopolysaccharide production by *Lactobacillus casei* CG11, isolated from cheese. *Appl. Environ. Microbiol.* 58, 4086–4088.
- Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. doi: 10.1093/bioinformatics/btm091
- Kralj, S., Van Geel-Schutten, G. H., Dondorff, M. M. G., Kirsanovs, S., Van Der Maarel, M. J. E. C., and Dijkhuizen, L. (2004a). Glucan synthesis in the genus *Lactobacillus*: isolation and characterization of glucanucrase genes, enzymes and glucan products from six different strains. *Microbiology* 150, 3681–3690. doi: 10.1099/mic.0.27321-0
- Kralj, S., Van Geel-Schutten, G. H., Van Der Maarel, M. J. E. C., and Dijkhuizen, L. (2003). Efficient screening methods for glucosyltransferase genes in *Lactobacillus* strains. *Biocatal. Biotransform.* 21, 181–187. doi: 10.1080/10242420310001618519
- Kralj, S., Van Geel-Schutten, G. H., Van Der Maarel, M. J. E. C., and Dijkhuizen, L. (2004b). Biochemical and molecular characterization of *Lactobacillus reuteri* 121 reuteranucrase. *Microbiology* 150(Pt 7), 2099–2112. doi: 10.1099/mic.0.27105-0
- Leemhuis, H., Pijning, T., Dobruchowska, J. M., van Leeuwen, S. S., Kralj, S., Dijkstra, B. W., et al. (2013). Glucanucleases: three-dimensional structures, reactions, mechanism,  $\alpha$ -glucan analysis and their implications in biotechnology and food applications. *J. Biotechnol.* 163, 250–272. doi: 10.1016/j.jbiotec.2012.06.037
- Macrina, F. L., Kopecko, D. J., Jones, K. R., Ayers, D. J., and McCowen, S. M. (1978). A multiple plasmid-containing *Escherichia coli* strain: convenient source of size reference plasmid molecules. *Plasmid* 1, 417–420. doi: 10.1016/0147-619X(78)90056-2
- Mohedano, M. L., García-Cayuela, T., Pérez-Ramos, A., Gaiser, R., Requena, T., and López, P. (2015). Construction and validation of a mCherry protein vector for promoter analysis in *Lactobacillus acidophilus*. *J. Ind. Microbiol. Biot.* 42, 247–253. doi: 10.1007/s10295-014-1567-4
- Monsan, P., Bozonnet, S., Albenne, C., Joucla, G., Willemot, R. M., and Remaud-Siméon, M. (2001). Homopolysaccharides from lactic acid bacteria. *Int. Dairy J.* 11, 675–685. doi: 10.1016/S0958-6946(01)00113-3
- Nácher-Vázquez, M., Ballesteros, N., Canales, A., Rodríguez Saint-Jean, S., Pérez-Prieto, S. I., Prieto, A., et al. (2015). Dextran produced by lactic acid bacteria exhibit antiviral and immunomodulatory activity against salmonid viruses. *Carbohydr. Polym.* 124, 292–301. doi: 10.1016/j.carbpol.2015.02.020
- Nácher-Vázquez, M., Iturria, I., Zarour, K., Mohedano, M. L., Aznar, R., Pardo, M. A., et al. (2017). Dextran production by *Lactobacillus sakei* MN1 coincides with reduced autoagglutination, biofilm formation and epithelial cell adhesion. *Carbohydr. Polym.* 168, 22–31. doi: 10.1016/j.carbpol.2017.03.024
- Naessens, M., Cerdobbel, A., Soetaert, W., and Vandamme, E. J. (2005). Review *Leuconostoc* dextranucrase and dextran: production properties and applications. *J. Chem. Tech. Biotechnol.* 80, 845–860. doi: 10.1002/jctb.1322
- Neely, W. B., and Nott, J. (1962). Dextranucrase, an induced enzyme from *Leuconostoc mesenteroides*. *Biochemistry* 1, 1136–1140. doi: 10.1021/bi00912a027
- Neubauer, H., Bauché, A., and Mollet, B. (2003). Molecular characterization and expression analysis of the dextranucrase DsrD of *Leuconostoc mesenteroides* Lcc4 in homologous and heterologous *Lactococcus lactis* cultures. *Microbiology* 149, 973–982. doi: 10.1099/mic.0.26029-0
- Pérez-Ramos, A., Nácher-Vázquez, M., Notararigo, S., López, P., and Mohedano, M. L. (2015). “Current and future applications of bacterial extracellular polysaccharides,” in *Probiotics, Prebiotics, and Synbiotics*, eds V. R. Preedy and R. R. Watson (Oxford: Elsevier), 329–344.
- Piret, J., Lamontagne, J., Bestman-Smith, J., Roy, S., Gourde, P., Désormeaux, A., et al. (2000). *In vitro* and *in vivo* evaluations of sodium lauryl sulfate and dextran sulfate as microbicides against herpes simplex and human immunodeficiency viruses. *J. Clin. Microbiol.* 38, 110–119.
- Providenti, M. A., O'Brien, J. M., Ewing, R. J., Paterson, E. S., and Smith, M. L. (2006). The copy-number of plasmids and other genetic elements can be determined by SYBR-Green-based quantitative real-time PCR. *J. Microbiol. Methods* 65, 476–487. doi: 10.1016/j.mimet.2005.09.007
- Quirasco, M., López-Munguía, A., Remaud-Simeon, M., Monsan, P., and Farrés, A. (1999). Induction and transcription studies of the dextranucrase gene in *Leuconostoc mesenteroides* NRRL B-512F. *Appl. Environ. Microbiol.* 65, 5504–5509.
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. doi: 10.1038/nprot.2010.5
- Rühmkorf, C., Bork, C., Mischnick, P., Rübsam, H., Becker, T., and Vogel, R. F. (2013). Identification of *Lactobacillus curvatus* TMW 1.624 dextranucrase and comparative characterization with *Lactobacillus reuteri* TMW 1.106 and *Lactobacillus animalis* TMW 1.971 dextranucleases. *Food Microbiol.* 34, 52–61. doi: 10.1016/j.fm.2012.11.002
- Sánchez, C., Neves, A. R., Cavalheiro, J., Moreira dos Santos, M., García-Quintáns, N., López, P., et al. (2008). Contribution of citrate metabolism to the growth of *Lactococcus lactis* CRL264 at low pH. *Appl. Environ. Microbiol.* 74, 1136–1144. doi: 10.1128/AEM.01061-07
- Schwab, C., Walter, J., Tannock, G. W., Vogel, R. F., and Gänzle, M. G. (2007). Sucrose utilization and impact of sucrose on glycosyltransferase expression in *Lactobacillus reuteri*. *Syst. Appl. Microbiol.* 30, 433–443. doi: 10.1016/j.syapm.2007.03.007
- Swe, P. M., Heng, N. C. K., Cook, G. M., Tagg, J. R., and Jack, R. W. (2010). Identification of DysI, the immunity factor of the streptococcal bacteriocin dysgalactin. *Appl. Environ. Microbiol.* 76, 7885–7889. doi: 10.1128/AEM.01707-10
- Untergasser, A., Cutcutache, L., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596



- van Hijum, S. A. F. T., Kralj, S., Ozimek, L. K., Dijkhuizen, L., and van Geel-Schutten, I. G. H. (2006). Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. *Microbiol. Mol. Biol. Rev.* 70, 157–176. doi: 10.1128/MMBR.70.1.157-176.2006
- Vujicic-Zagar, A., Pijning, T., Kralj, S., Lopez, C. A., Eeuwema, W., Dijkhuizen, L., et al. (2010). Crystal structure of a 117 kDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21406–21411. doi: 10.1073/pnas.1007531107
- Wang, T. T., and Lee, B. H. (1997). Plasmids in *Lactobacillus*. *Crit. Rev. Biotechnol.* 17, 227–272. doi: 10.3109/07388559709146615
- Wenham, D. G., Hennessey, T. D., and Cole, J. A. (1979). Regulation of lucosyl- and fructosyltransferase synthesis by continuous cultures of *Streptococcus mutans*. *J. Gen. Microbiol.* 114, 117–124. doi: 10.1099/00221287-114-1-117
- Werning, M. L., Ibarburu, I., Due-as, M. T., Irastorza, A., Navas, J., and López, P. (2006). *Pediococcus parvulus* gtf gene encoding the GTF glycosyltransferase and Its application for specific PCR detection of beta-D-glucan-producing bacteria in foods and beverages. *J. Food Prot.* 69, 161–169. doi: 10.4315/0362-028X-69.1.161
- Werning, M. L., Notararigo, S., Nácher, M., Fernández de Palencia, P., Aznar, R., and López, P. (2012). “Biosynthesis, purification and biotechnological use of exopolysaccharides produced by lactic acid bacteria,” in *Food Additives*, ed Y. El-Samragy (Rijeka: Intech), 83–114.
- Yanping, Y., Peng, Q., Guo, Y., Han, Y., Xiao, H., and Zhou, Z. (2015). Isolation and characterization of dextran produced by *Leuconostoc citreum* NM105 from manchurian sauerkraut. *Carbohydr. Polym.* 133, 365–372. doi: 10.1016/j.carbpol.2015.07.061
- Zarour, K., Llamas, M. G., Prieto, A., Rúas-Madiedo, P., Due-as, M. T., de Palencia, P. F., et al. (2017). Rheology and bioactivity of high molecular weight dextrans synthesised by lactic acid bacteria. *Carbohydr. Polym.* 174, 646–657. doi: 10.1016/j.carbpol.2017.06.113

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Nácher-Vázquez, Ruiz-Masó, Mohedano, del Solar, Aznar and López. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Plasmid-Mediated Bioaugmentation for the Bioremediation of Contaminated Soils

Carlos Garbisu<sup>1</sup>, Olatz Garaiyurrebaso<sup>2</sup>, Lur Epelde<sup>1</sup>, Elisabeth Grohmann<sup>3</sup> and Itziar Alkorta<sup>1\*</sup>

<sup>1</sup> Soil Microbial Ecology Group, Department of Conservation of Natural Resources, Neiker Tecnalia, Derio, Spain, <sup>2</sup> Instituto Biofisika (UPV/EHU, CSIC), Department of Biochemistry and Molecular Biology, University of the Basque Country, Bilbao, Spain, <sup>3</sup> Beuth University of Applied Sciences, Berlin, Germany

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
United States

### Reviewed by:

Gloria Del Solar,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain  
Spiros Nicolas Agathos,  
Catholic University of Louvain,  
Belgium  
Fabián Lorenzo,  
Universidad de La Laguna, Spain

### \*Correspondence:

Itziar Alkorta  
itzi.alkorta@ehu.eus

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 20 May 2017

Accepted: 25 September 2017

Published: 09 October 2017

### Citation:

Garbisu C, Garaiyurrebaso O,  
Epelde L, Grohmann E and Alkorta I  
(2017) Plasmid-Mediated  
Bioaugmentation  
for the Bioremediation  
of Contaminated Soils.  
Front. Microbiol. 8:1966.  
doi: 10.3389/fmicb.2017.01966

Bioaugmentation, or the inoculation of microorganisms (e.g., bacteria harboring the required catabolic genes) into soil to enhance the rate of contaminant degradation, has great potential for the bioremediation of soils contaminated with organic compounds. Regrettably, cell bioaugmentation frequently turns into an unsuccessful initiative, owing to the rapid decrease of bacterial viability and abundance after inoculation, as well as the limited dispersal of the inoculated bacteria in the soil matrix. Genes that encode the degradation of organic compounds are often located on plasmids and, consequently, they can be spread by horizontal gene transfer into well-established, ecologically competitive, indigenous bacterial populations. Plasmid-mediated bioaugmentation aims to stimulate the spread of contaminant degradation genes among indigenous soil bacteria by the introduction of plasmids, located in donor cells, harboring such genes. But the acquisition of plasmids by recipient cells can affect the host's fitness, a crucial aspect for the success of plasmid-mediated bioaugmentation. Besides, environmental factors (e.g., soil moisture, temperature, organic matter content) can play important roles for the transfer efficiency of catabolic plasmids, the expression of horizontally acquired genes and, finally, the contaminant degradation activity. For plasmid-mediated bioaugmentation to be reproducible, much more research is needed for a better selection of donor bacterial strains and accompanying plasmids, together with an in-depth understanding of indigenous soil bacterial populations and the environmental conditions that affect plasmid acquisition and the expression and functioning of the catabolic genes of interest.

**Keywords:** biodegradation, catabolic plasmid, fitness cost, horizontal gene transfer, soil pollution

## INTRODUCTION

Soils play a vital role in the provision of ecosystem services and harbor one of the most complex and diverse biological communities on Earth (Barrios, 2007). Therefore, the preservation of soil quality/soil health (both terms are often used interchangeably), defined as “the capacity of soil to perform its ecosystem processes and services, while maintaining ecosystem attributes of ecological relevance” (Garbisu et al., 2011), is currently a matter of great priority. Contamination is one of the most important causes of soil degradation. Only in Europe, there are around 2.5 million potentially

contaminated sites, with an annual estimated management cost of 6 billion euros (Panagos et al., 2013). Different anthropogenic activities, such as combustion of fossil fuels, incineration, mining, agricultural practices, urbanization, waste disposal, etc. have contributed to the pressing problem of soil contamination (Besser et al., 2009). Among other negative consequences, the presence of contaminants in soil can cause a negative impact on the soil biota, resulting in an altered activity, biomass and/or diversity of soil biological communities (Burgess et al., 2015).

## REMEDIATION OF SOIL CONTAMINANTS

Traditionally, a variety of physicochemical methods (e.g., excavation and disposal in landfills, soil washing, chemical oxidation, encapsulation, thermal treatments, incineration, vitrification, solidification, etc.) have been used for soil remediation. However, these physicochemical strategies are often expensive and, many times, reduce the concentration of soil contaminants at the expense of damaging the integrity of the soil ecosystem (Epelde et al., 2009; Gómez-Sagasti et al., 2016).

The main goal of any soil remediation technology must be not only to reduce the concentration of soil contaminants but to restore soil quality (Epelde et al., 2010; Barrutia et al., 2011; Pardo et al., 2014). A variety of soil physicochemical and biological properties (e.g., parameters that provide information on the biomass, activity and diversity of soil microbial communities) (Epelde et al., 2009; Muñoz-Leoz et al., 2013) are often used as indicators of soil quality. It has also been proposed to assess the effectiveness of remediation methods in terms of the recovery of soil ecosystem services and/or attributes of ecological relevance, such as organization, stability, redundancy, etc. (Garbisu et al., 2011; Epelde et al., 2014).

As an alternative to physicochemical treatments, several biological methods of soil remediation, included within the terms bioremediation and phytoremediation, are currently receiving much attention, mainly owing to their lower cost and environmentally friendly character (Juwarkar et al., 2014). Bioremediation, or the use of microorganisms (bacteria, fungi) to break down contaminants, takes advantage of the catabolic capacity of microorganisms to remove contaminants from soil. However, bioremediation is effective only with a limited range of contaminants and contaminant concentrations. In addition, bioremediation techniques might take too long to achieve the desired reduction in the concentration of soil contaminants (Kumavath and Deverapalli, 2013).

In relation to trace elements (a group of non-degradable contaminants of much concern due to their well-known toxicity), microorganisms can only transform them from one oxidation state or organic complex to another (Garbisu et al., 2002). Then, for the biological remediation of metal contaminated soils, metal-accumulating plants (i.e., accumulators and hyperaccumulators) offer many advantages over microbial processes, as these plants can literally extract the toxic metals from the contaminated site through a phytotechnology termed phytoextraction (Barrutia et al., 2009, 2010; Epelde et al., 2010).

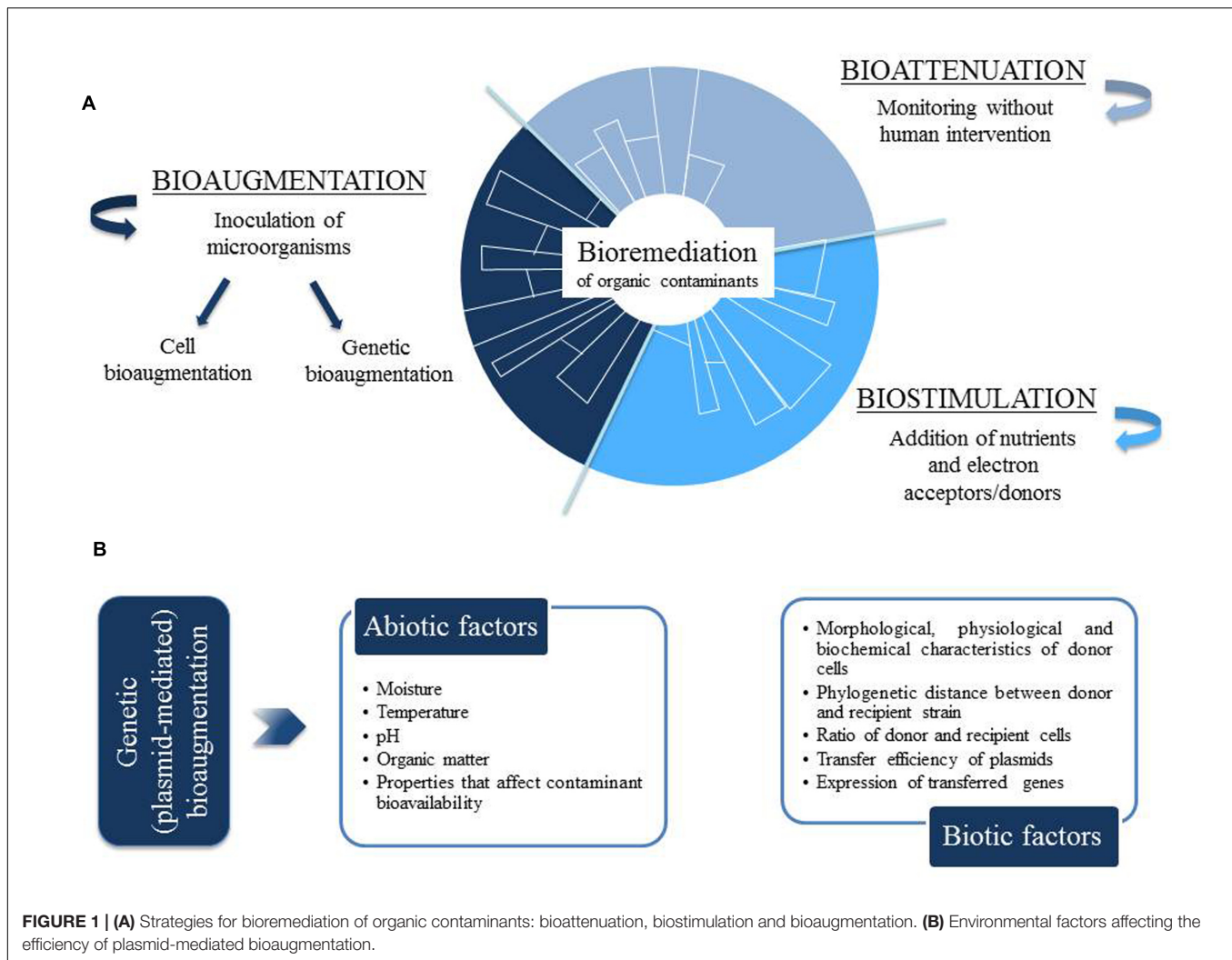
Bioremediation has been successfully employed to remediate soils contaminated with organic contaminants, such as aliphatic hydrocarbons, polycyclic aromatic hydrocarbons, polychlorinated biphenyls, organic solvents and so on (Maphosa et al., 2012).

The bioremediation of organic contaminants can be approached by three different strategies: bioattenuation, biostimulation, and bioaugmentation (Figure 1A). *Bioattenuation* relies on natural processes to maintain the growth and degrading activity of native microbial populations, so that contaminants are biodegraded without human intervention, apart from the monitoring of contaminant dispersal and degradation rates. Instead, the term *biostimulation* refers to the adjustment of the environmental conditions (e.g., temperature, moisture, aeration, pH, redox potential) and the application of nutrients (e.g., nitrogen, phosphorus) and electron acceptors to contaminated soil, in order to enhance the growth of degrading microbial populations and, then, reduce the concentration of soil contaminants. Finally, *bioaugmentation* has been defined as the inoculation into contaminated soils of microorganisms with the ability to degrade the target contaminants (Maier, 2000; Heinaru et al., 2005). This inoculation can be performed with only one strain or, alternatively, with a consortium of microbial strains with diverse metabolic capacities. The advantage of using a consortium of different strains is that toxic intermediate products generated by one strain may be degraded by another strain (Heinaru et al., 2005). Apart from inoculating wild strains with the required degradation capacities, laboratory-constructed strains with upgraded catabolic abilities have also been considered for a more efficient bioaugmentation (Mrozik et al., 2011).

Iwamoto and Nasu (2001) and El Fantroussi and Agathos (2005) have proposed to apply bioaugmentation in those cases where biostimulation and natural attenuation are proven ineffective. In this regard, in a diesel-contaminated soil, Bento et al. (2005) found bioaugmentation to be more effective for the degradation of the light fraction (C12–C23) of petroleum hydrocarbons than biostimulation. No significant differences were detected between biostimulation and bioaugmentation in relation to the removal of the heavy fraction (C23–C40).

Bioaugmentation can be divided into two different approaches: (i) *cell bioaugmentation*, which relies on the survival and growth of the inoculated strains to perform the degradation of the target contaminants, and (ii) *genetic bioaugmentation*, based on the spread of catabolic genes, located in mobile genetic elements (MGEs), into native microbial populations.

However, despite decades of bioremediation research, the real drivers governing the degradation of organic contaminants are still poorly understood (Meckenstock et al., 2015). In order to gain insight into this question, Meckenstock et al. (2015) revisited and challenged current concepts on the controls and limitations of biodegradation, and pointed out some critical research gaps such as, for instance, the role of protozoa and bacteriophages in shaping communities of bacterial degraders and influencing contaminant degradation rates.



## CELL BIOAUGMENTATION

Cell bioaugmentation is based on the survival and catabolic activity of inoculated microbial strains (Singh and Ward, 2004). The inoculation of bacteria harboring the necessary metabolic pathways for the degradation of the target contaminants can indeed accelerate the removal of such contaminants and, hence, reduce the time required for the intended bioremediation (Nowak and Mroziak, 2016). Inoculated microbial strains must then compete for energy and resources (e.g., nutrients and electron acceptors) with the autochthonous microbial populations already present in the soil ecosystem. The major drawbacks for the successful application of cell bioaugmentation are the (i) frequently very high mortality of the inoculated microbial strains, due to biotic or abiotic stresses, and (ii) limited dispersal of such strains throughout the soil matrix (Pepper et al., 2002; Quan et al., 2010). Many factors, including cell adhesion to soil organic matter (OM), can strongly limit the distribution of bacteria through the soil matrix. To overcome this limitation, several authors

(Wang and Mulligan, 2004; Franzetti et al., 2009) have reported the use of surfactants, foams and adhesion-resistant strains.

Despite these limitations, many studies have supported the potential of cell bioaugmentation for the bioremediation of soils contaminated with organic compounds. Wang et al. (2004) reported an accelerated removal of quinoline after the inoculation of *Burkholderia pickettii*. Similarly, Mroziak et al. (2011) showed that cell bioaugmentation with *Pseudomonas* sp. JS150 significantly enhanced phenol degradation in soil, thereby reducing the possibility of formation of phenoxyl radicals (Hansch et al., 2000). Although the number of *Pseudomonas* sp. JS150 cells decreased significantly during the first few days, the inoculated bacteria were then able to survive over the experimental period and successfully increased the rate of phenol degradation; actually, phenol biodegradation in soil bioaugmented with *Pseudomonas* sp. JS150 cells was 68 and 96 days shorter in clay and sandy soil, respectively, in comparison to non-bioaugmented soil (Mroziak et al., 2011).



## GENETIC (PLASMID-MEDIATED) BIOAUGMENTATION

Genes encoding the degradation of naturally occurring or xenobiotic organic compounds are often located on MGEs, such as plasmids, integrons and transposons. By acquiring these genes through mechanisms of horizontal gene transfer (HGT), recipient bacteria may achieve the capacity to degrade those organic contaminants (Wiedenbeck and Cohan, 2011). HGT allows the exchange of genetic information among bacteria from even distantly related taxonomic groups, thereby allowing bacteria to rapidly adapt to new environmental conditions. Although mutation events can certainly contribute to bacterial adaptation, mutation rates in bacterial populations are generally low. Besides, it is currently assumed that an increased rate of mutations would result in increased death owing to deleterious effects (Martínez et al., 2009).

Out of the three mechanisms of HGT in bacteria (i.e., transformation, transduction and conjugation), conjugation is a most efficient biological process in which genetic information encoded in plasmids is transferred, from donor to recipient bacteria, by direct cell-to-cell contact (Furuya and Lowy, 2006). Bacterial conjugation is known to accelerate the dissemination of resistance to, for instance, antibiotics and heavy metals, as well as to facilitate the distribution of genes involved in the degradation of organic compounds. Nevertheless, the contribution of conjugation to HGT among soil bacteria and the factors involved in the transfer and proliferation of plasmid-containing bacteria in the soil ecosystem are yet not fully understood.

Bacterial adaptation through evolutionary time has been shaped, among other aspects, by the high plasticity of bacterial genomes, which allows bacteria to rearrange and exchange genomic sequences, thus opening the possibility to acquire beneficial traits (Sørensen et al., 2005). As a matter of fact, the loss, rearrangement and acquisition of functional genetic modules can have a vast impact on the extent and speed of the evolutionary adaptation of bacteria (Wozniak and Waldor, 2010; Bertels and Rainey, 2011). MGEs are, to a great extent, responsible for these processes of gene mobility and reorganization, both within genomes (intracellular) and between bacterial cells (intercellular).

Many of the studies on lateral dissemination of genetic material among bacteria have focused on antibiotic and metal resistance. Research on the horizontal transfer of genes associated with the degradation of organic compounds in natural environments, such as the soil ecosystem, is still insufficient to fully understand the mechanisms involved in such process (Christensen et al., 1998; Top et al., 1998; Dejonghe et al., 2000; Aspray et al., 2005; Overhage et al., 2005; Musovic et al., 2010). In any case, some plasmids, such as those implicated in the catabolic pathway of 2,4-dichlorophenoxyacetic acid (2,4-D), have been thoroughly studied (Top et al., 1998; Dejonghe et al., 2000; Newby and Pepper, 2002).

Plasmid transfer between soil bacteria has been contemplated as a promising strategy for the dissemination of catabolic functions within soil bacterial communities (Venkata Mohan et al., 2009; Mroziak and Piotrowska-Seget, 2010).

As abovementioned, plasmid-encoded metabolic pathways can be transferred among bacteria, thus playing a critical role in the adaptation of bacteria to different environmental conditions (Reineke, 1998; Sayler and Ripp, 2000). Specifically, HGT has been reported to promote bacterial adaptation to the presence of organic contaminants (Top and Springael, 2003).

The underlying idea behind genetic (plasmid-mediated) bioaugmentation is to stimulate the rate of contaminant degradation by increasing, through HGT, the number and diversity of native bacteria with the capacity to metabolize the target contaminants. In this respect, it must be emphasized that numerous catabolic pathways involved in the degradation of organic contaminants have been identified in MGEs (Top et al., 2002; Jussila et al., 2007).

Genetic (plasmid-mediated) bioaugmentation is defined as a technology in which donor bacteria harboring self-transmissible catabolic plasmids are introduced into the soil matrix in order to enhance, by HGT, the potential and rate of contaminant degradation of existing bacterial populations (Top et al., 2002; Ikuma and Gunsch, 2010, 2012). Compared to cell bioaugmentation, plasmid-mediated bioaugmentation appears *a priori* a more effective strategy for the bioremediation of organic contaminants, as the bacteria that will eventually degrade the contaminants (i.e., bacteria with the recently acquired plasmids harboring the necessary catabolic genes) are expected to be adapted to live in the soil under remediation. In this manner, one of the main drawbacks for the successful application of cell bioaugmentation, i.e., the low survival of the inoculated microbial strains, appears to be overcome.

For plasmid-mediated bioaugmentation, both an appropriate selection of donor bacteria with the required plasmid and a profound knowledge of native soil bacterial populations are required to increase the probability of an efficient plasmid acquisition and the expression of the catabolic genes of interest.

Many studies on plasmid-mediated bioaugmentation have been published (Table 1). In a microcosm study, Halden et al. (1999) detected an enhanced degradation of 3-phenoxybenzoic acid (3-POB) as a result of the transfer of plasmids pPOB and pD30.9 from *Pseudomonas pseudoalcaligenes* POB310 (pPOB) and *Pseudomonas* sp. B13-D5 (pD30.9) and B13ST1 (pPOB) to recipient soil bacteria. Using *P. putida* as donor strain of two catabolic plasmids (pEMT1 and pJP4), Dejonghe et al. (2000) reported the degradation of 2,4-D in soil under microcosm conditions. These authors investigated the bioaugmentation potential of plasmids pEMT1 and pJP4 in two soil layers (0–30 and 30–60 cm soil depth) differing in physicochemical properties and microbial community structure, finding out a more efficient degradation of 2,4-D in the deeper soil layer where the indigenous microbial communities lacked the ability to catabolize 2,4-D. Under microcosm conditions, Inoue et al. (2012) studied the effect of bioaugmentation with *P. putida* and *Escherichia coli* cells, harboring the self-transmissible 2,4-D degradative plasmid pJP4, on the degradation of 2,4-D. These authors found that the number of *P. putida* and *E. coli* cells decreased rapidly after their inoculation in a 2,4-D contaminated soil slurry, but the degradation of this contaminant was nevertheless stimulated, most likely due to the occurrence of

transconjugants resulting from the transfer of plasmid pJP4. Inoue et al. (2012) concluded that genetic bioaugmentation with *P. putida* and *E. coli* cells harboring plasmid pJP4 can stimulate the degradation of 2,4-D in soil without a substantial impact on the soil microbial community, as reflected by the values of parameters which provide information on carbon source utilization (through the use of the well-known Biolog<sup>TM</sup> plates) and nitrogen transformations (nitrate reduction assay, quantification of amoA gene of ammonia-oxidizing bacteria, quantification of *nirK* and *nirS* genes of denitrifying bacteria). In sequencing batch reactors, Tsutsui et al. (2013) achieved a complete degradation of 2,4-D by plasmid (pJP4)-mediated bioaugmentation with *Cupriavidus necator* JMP134 and *E. coli* HB101 as donor strains. These authors were able to identify the emergence of 2,4-D-degrading transconjugants associated to *Achromobacter*, *Burkholderia*, *Cupriavidus* and *Pandoraea*.

Pepper et al. (2002) conducted microcosm experiments to enhance the degradation of 3-chlorobenzoate (3-CB) using plasmid pBRC60, which harbors genes for 3-CB mineralization, and *Comamonas testosteroni* as donor strain. Although they did observe degradation of 3-CB, they could not detect any transfer event of plasmid pBRC60 from *C. testosteroni* to native soil bacteria.

Miyazaki et al. (2006) isolated a plasmid (pLB1) involved in the dissemination of genes for  $\gamma$ -hexachlorocyclohexane (lindane) degradation. This plasmid, carrying the *linB* gene, was isolated from *Sphingobium japonicum* UT26DB and then successfully transferred, under laboratory conditions, from this strain to other  $\alpha$ -proteobacterial strains but not to any of the  $\beta$ - or  $\gamma$ -proteobacterial strains tested.

In their study on the transfer of TOL plasmid (also designated pWW0) during bacterial conjugation *in vitro* and rhizoremediation of oil-contaminated soil *in vivo*, Jussila et al. (2007) demonstrated the successful transfer of TOL plasmid for toluene degradation from *P. putida* PaW85 to *P. oryzae* 29. In rhizosphere microcosms, Mølbak et al. (2007) found that the transfer of plasmid pWW0 from *P. putida* resulted in transconjugants belonging to *Enterobacteria* and *Pseudomonas*. This well-characterized self-transmissible catabolic plasmid, pWW0, was also used by Ikuma and Gunsch (2012) to assess

its potential for bioaugmentation in toluene-contaminated soil slurry.

Under laboratory conditions, horizontal transfer of plasmid pGKT2 was successfully carried out by Jung et al. (2011) from *Gordonia* sp. KTR9 to *Gordonia polyisoprenivorans*, *Rhodococcus jostii* RHA1 and *Nocardia* sp. TW2 strains. These transconjugants showed the ability to use hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) as a nitrogen source.

In a contaminated field site located in Cixi, Zhejiang (China), Gao et al. (2015) achieved effective plasmid-mediated bioaugmentation for the degradation of dichlorodiphenyltrichloroethane (DDT) in soil with *E. coli* TG I (pDOD-gfp) as donor strain. In this study, the catabolic plasmid pDOD from *Sphingobacterium* sp. D-6 was conjugally transferred to soil bacteria, such as members of *Cellulomonas*, and accelerated DDT degradation. Different studies have reported the use of the GFP (green fluorescence protein) detection system to monitor plasmid transfer from donor cells to indigenous soil bacteria in soil slurries (Ikuma et al., 2012) and field contaminated soil (Gao et al., 2015).

Filonov et al. (2010) determined transfer frequencies in open soil after inoculation with genetically tagged plasmid-containing naphthalene-degrading *P. putida* KT2442 and auxotrophic donor BS394 (pNF142::TnMod-OTc) cells, finding out that plasmid pNF142 was transferred to native soil bacteria (mainly to fluorescent pseudomonads) at a frequency of  $4 \times 10^{-6}$ /donor cell. After bioaugmentation with *E. coli* JM109 (pDOC-gfp) strain, Zhang et al. (2012) observed that pDOC plasmid was transferred to native soil bacteria under microcosm conditions, including members of *Pseudomonas* and *Staphylococcus* which acquired the capacity to degrade chlorpyrifos (a widely used insecticide). As it is usually the case, the efficiency of this transfer, as measured by the chlorpyrifos degradation efficiency and the number of chlorpyrifos degraders, was influenced by soil type, temperature and moisture content (Zhang et al., 2012).

Finally, attention has also been paid to the use of genetically modified organisms (GMOs) for bioaugmentation. Nonetheless, the deliberate release of GMOs into the environment is subjected to regulatory constraints (Garbisu and Alkorta, 1997; Sayler and Ripp, 2000; Directive 2001/18/EC, 2001). The transfer of

**TABLE 1** | Examples of plasmid-mediated bioaugmentation studies.

Contaminant	Plasmid	Strain	Reference
3-phenoxybenzoic acid	pPOB	<i>P. pseudoalcaligenes</i> POB310	Halden et al., 1999
3-chlorobenzoate	pBRC60	<i>C. testosteroni</i>	Pepper et al., 2002
2,4-dichlorophenoxyacetic acid	pJP4	<i>C. necator</i> JMP134 <i>E. coli</i> HB101	Tsutsui et al., 2013
2,4-dichlorophenoxyacetic acid	pJP4	<i>P. putida</i> KT2442 <i>E. coli</i> HB101	Inoue et al., 2012
2,4-dichlorophenoxyacetic acid	pEMT1 pJP4	<i>P. putida</i> UWC3	Dejonghe et al., 2000
2,4-dichlorophenoxyacetic acid	pJP4	<i>R. eutropha</i> JMP134 <i>E. coli</i> D11	Newby et al., 2000
$\gamma$ -hexachlorocyclohexane	pLB1	<i>S. japonicum</i> UT26DB	Miyazaki et al., 2006
Chlorpyrifos	pDOC	<i>E. coli</i> JM109	Zhang et al., 2012
Dichlorodiphenyltrichloroethane	pDOD	<i>E. coli</i> TG I	Gao et al., 2015
Naphthalene	pNF142	<i>P. putida</i> BS394 <i>P. putida</i> KT2442	Filonov et al., 2010
Oil	pWW0-derivative TOL	<i>P. putida</i> PaW85	Jussila et al., 2007
Toluene	pWW0-derivative TOL	<i>P. putida</i> BBC443	Ikuma and Gunsch, 2012

catabolic genes between GMOs and wild bacterial strains might facilitate the acquisition and spread of new degradative pathways among indigenous bacterial communities. To this purpose, Massa et al. (2009) engineered the recombinant strain *P. putida* PaW340/pDH5, constructed by cloning dehalogenase genes from *Arthrobacter* sp. FG1 in *P. putida* PaW340, for the degradation of 4-chlorobenzoic acid (CBA) in soil slurry. After inoculation of this recombinant strain into soil slurry, a higher degradation of CBA was observed, compared to the slurry inoculated with pre-adapted cultures of *Arthrobacter* sp. FG1.

## EFFECT OF PLASMID ACQUISITION ON HOST FITNESS

The success of plasmid-mediated bioaugmentation for the bioremediation of contaminated soil relies not only on an efficient transfer of the required plasmid from donor bacteria to soil recipient bacteria, but also on the ability of recipient cells to properly express the plasmid-harbored catabolic genes, so that the desired phenotypic changes (i.e., biodegradation of the target contaminant) can be attained. After plasmid acquisition, the capacity of recipient cells to successfully perform the desired catabolic function depends, among other factors, on their newly acquired competitive abilities and on the alteration of the host's own competitive abilities (van Rensburg et al., 2012). A thorough understanding of how plasmid acquisition can affect host fitness is fundamental to then achieve the persistence of the introduced plasmid in the recipient cells.

Plasmid acquisition can provide recipient bacteria with a large array of beneficial traits, such as catabolic potential, resistance to antibiotics and/or metals, faster growth, ability to use a wider range of compounds as energy sources, etc. (Top et al., 1998; Riley and Wertz, 2002). In many cases, plasmid-harboring hosts have been found to be competitively fitter than their plasmid-free counterparts (Dionisio et al., 2005; Starikova et al., 2013). Nonetheless, horizontally acquired genes can also function inefficiently in the genomic background of recipient cells (Chou et al., 2011; Park and Zhang, 2012). After all, horizontally acquired genes find themselves immersed in a new metabolic context and their function relies on the host's machinery. Genetic determinants often encounter the required metabolic "partners" (e.g., substrates, proteins) in the recipient cells, so that the intended changes in the host's metabolism become possible. Conversely, other times, the required metabolic partners for the proper functioning and regulation of newly acquired genes are missing in the recipient cells. Indeed, the acquisition of plasmids can negatively affect cellular networks in recipient cells and, concomitantly, trigger fitness costs as collateral damage (Bouma and Lenski, 1988; Martínez et al., 2009). Fitness (metabolic) costs derived from plasmid acquisition can be highly variable (De Gelder et al., 2007), as they can originate from a variety of factors, including: (i) *energetic costs* due to consumption of molecular building blocks and/or energy sources derived from the activity of horizontally acquired regions; (ii) *chromosomal disruption* by horizontally acquired genes, when such genes are incorporated into the chromosome; (iii) *sequestration of cellular*

*processes and associated molecular machinery* (e.g., ribosomes) by the horizontally acquired regions; and (iv) *plasmid size*, since small plasmids can carry only a single accessory determinant but large plasmids can carry more than 10 accessory determinants as well as other genes (Shachrai et al., 2010; Baltrus, 2013; Vogwill and MacLean, 2015).

Fitness costs associated to plasmid acquisition can be offset by benefits derived from the fact that plasmids are ideal biological tools to create genetic variation within bacterial populations. A major benefit from maintaining transferable plasmids derives from the fact that, in this manner, bacterial populations can gain stability against potential environmental changes.

Bacteria with acquired genes can, on the other hand, alleviate fitness costs through compensatory evolution (San Millan et al., 2014). Thus, for instance, bacteria can minimize plasmid-related fitness costs by integrating only the desired plasmid-acquired determinants in the chromosome.

Conjugative plasmids (i) are usually large (they encode genes for the conjugation process itself and for stabilization within the host); (ii) are normally found in low copy number; (iii) appear well maintained over successive generations (Norman et al., 2009; Jung et al., 2011); and (iv) act as fundamental vehicles of HGT (Frost et al., 2005; Thomas and Nielsen, 2005).

In their laboratory study on the capacity of *Gordonia* sp. KTR9 to transfer plasmid pGKT2 and the associated RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine) degradation ability to other bacteria, Jung et al. (2011) investigated plasmid stability after HGT from *Gordonia* sp. KTR9 to *G. polyisoprenivorans*, *R. jostii* RHA1 and *Nocardia* sp. TW2, finding out a marked decrease in plasmid retention after 50 generations with *Nocardia* sp. TW2, while *G. polyisoprenivorans* and *R. jostii* RHA1 transconjugants exhibited retention of pGKT2 plasmid for 100 generations. It was speculated that this decreased stability in *Nocardia* sp. TW2 might have been caused by a larger metabolic expense incurred by the incorporation of pGKT2 in this strain, compared to the other two bacterial strains (Jung et al., 2011).

Given that positive selection cannot explain the long-term stability of costly plasmids (Hall et al., 2017), the explanation for such long-term stability remains a most challenging task, since segregational loss and the cost of plasmid carriage should drive the loss of plasmids through purifying selection (Hall et al., 2017). In this respect, two evolutionary routes to plasmid stability appear possible (Hall et al., 2017): (i) the evolution of high conjugation rates would allow plasmids to survive as infectious agents through horizontal transmission (Hall et al., 2016; Kottara et al., 2016); and (ii) compensatory evolution to ameliorate the cost of plasmid carriage can weaken purifying selection against the plasmid backbone (Harrison et al., 2015; Porse et al., 2016).

Finally, it must be taken into consideration that plasmids can be classified into incompatibility groups (incompatibility defined as the inability of plasmids sharing similar replication and partition systems to be propagated stably in the same host cell line; in other words, members of each group cannot co-reside within the same bacterial host), such as IncP, IncN, IncW, and IncF. Incompatibility groups have been independently classified in three different genera: there



are 27 Inc groups in *Enterobacteriaceae*, 14 Inc groups in *Pseudomonas*, and approximately 18 Inc groups in *Staphylococcus* (Shintani et al., 2015). Plasmids classified in *E. coli* as IncP and in *Pseudomonas* as IncP-1 are a well-studied group of plasmids that can carry a variety of phenotypic markers, including antibiotic resistance, metal resistance and the ability to degrade xenobiotics. It has been reported (Popowska and Krawczyk-Balska, 2013) that a detailed analysis of IncP-1 plasmid genomes could provide useful information for the development of effective methods of soil bioremediation. After all, the evolutionary adaptation of microorganisms to the presence and utilization of organic contaminants is often due to plasmids (mainly, from IncP group) that carry genes encoding enzymes involved in the degradation of those compounds. For instance, plasmids IncP-1, IncP-7 and IncP-9 contain genes encoding enzymes required for the degradation of naphthalene, toluene, chlorobenzene, p-toluenesulfonate, 2,4-D, haloacetate and atrazine (Shintani et al., 2010a,b; Popowska and Krawczyk-Balska, 2013). Relevantly, there seems to be a distinction between (i) plasmids that harbor genes for the degradation of naturally occurring compounds and (ii) plasmids that harbor genes for the degradation of xenobiotics (Top et al., 2002): degradation of naturally occurring compounds is often encoded in IncP-2 and IncP-9 plasmids, while the degradation of xenobiotics seems to be encoded by the well-known broad host range IncP-1 plasmids. IncP-1 plasmids are very promiscuous, and this promiscuity appears to play a crucial role in the evolution of new metabolic pathways by recruiting catabolic genes or gene segments from different organisms into a suitable host (Wyndham et al., 1994; Beil et al., 1999).

Therefore, different plasmids potentially useful for plasmid-mediated bioaugmentation (with, for instance, each plasmid harboring a gene encoding a different enzyme involved in the degradation route of a specific contaminant) cannot co-reside within the same host if they belong to the same incompatibility group. Then, if we want to apply different plasmids from the same incompatibility group, each of them harboring a gene for a specific step in the contaminant degradation pathway, they must be applied in different donor cells and, for an effective biodegradation, each plasmid should be transferred to a different recipient cell, decreasing considerably the probability of successful plasmid-mediated bioaugmentation.

## INFLUENCE OF ABIOTIC AND BIOTIC FACTORS ON BIOAUGMENTATION

The success of both cell and plasmid-mediated bioaugmentation greatly depends on the environmental (abiotic and biotic) conditions present in the soil to be remediated (Cho et al., 2000; Bento et al., 2005; Wolski et al., 2006). In fact, during plasmid-mediated bioaugmentation, environmental factors can play important roles in the (i) transfer efficiency of catabolic plasmids, (ii) expression of horizontally acquired genes and, finally, (iii) contaminant degradation activity (Popa et al., 2011; Ikuma and Gunsch, 2012). In particular, several abiotic factors

such as soil moisture, temperature and OM content are known to affect bioaugmentation efficiency (Figure 1B).

Soil moisture can have an effect on plasmid transfer during plasmid-mediated bioaugmentation by affecting the contact between donor and recipient bacteria (Miller et al., 2004; Aminov, 2011). In this respect, Gao et al. (2015) evaluated the effectiveness of plasmid-mediated bioaugmentation for *p,p'*-DDT degradation at three different soil moisture conditions (40, 60, and 80%), and concluded that 60% moisture content was optimal for maximum plasmid transfer efficiency.

Temperature has been shown to affect plasmid transfer efficiency (Inoue et al., 2005; Zhang et al., 2012). For the enhancement of DDT degradation by plasmid-mediated bioaugmentation with plasmid pDOD, the optimal temperature interval for cell growth and activity of both donor and recipient soil bacteria was established at 25–30°C (Gao et al., 2015). Johnsen and Kroer (2007) found that increasing temperatures resulted in an increase in the transfer of plasmid pRO103 encoding resistance to mercury and tetracycline and partial degradation of 2,4-D.

Regarding soil OM content, in a bioaugmentation laboratory experiment, Greer and Shelton (1992) observed higher rates of mineralization of 2,4-D in soil with a low OM content, compared to soil with a high content of OM. Under laboratory conditions, Kim et al. (2008) found that *P. spadix* BD-a59 cells were able to degrade BTEX at a slower rate in soil with low OM content than in organic-rich soil. When studying the biodegradation of polychlorinated biphenyls (PCB) in soil under laboratory conditions, Haluška et al. (1995) observed that humic acids affected the survival and activity of the inoculated *Alcaligenes xylosoxidans* strain, which exhibited maximum survival rates in soil with an intermediate amount of organic carbon and the highest amount of aromatic carbon in humic acids. Highest levels of PCB degradation were found in soil with the highest content of organic carbon and an intermediate amount of aromatic carbon in humic acids (Haluška et al., 1995).

Wang et al. (2014) performed plasmid transfer experiments between soil bacteria, using a TOL-like plasmid carrying the gene encoding for catechol 2,3-dioxygenase, to study some factors (soil depth, soil type, etc.) that could affect the transfer of plasmids, finding out that these factors certainly have a considerable effect on the transfer of the TOL-like plasmid in soil. Concerning soil depth, under microcosm conditions, Wang et al. (2014) found, in general, lower frequencies of plasmid transfer at greater soil depths, a fact most likely due to the often-found gradual decrease in bacterial biomass and activity at increasing soil depths, possibly related to concomitantly decreased oxygen concentrations (Król et al., 2011). Król et al. (2011) reported that oxygen concentration can affect plasmid transfer through an oxygen-related mechanism or indirectly via its impact on cell physiology. When studying the influence of soil type (loamy sand, sandy loam, sandy clay loam, loam) on plasmid transfer, Wang et al. (2014) observed a highest frequency of plasmid transfer in loam soil, probably related to the fact that loam often contains more nutrients and humus than other soil types, and higher values of microbial biomass and metabolic activity (Djokic et al., 2013).



In the same way, the chemical nature, concentration and bioavailability of the contaminants are crucial factors influencing bioaugmentation efficiency (Davis and Madsen, 1996; Stalwood et al., 2005). Sejáková et al. (2009) reported a relationship between pentachlorophenol (PCP) concentration in soil and the number of CFU of the *C. testosteroni* CCM7530 strain used for bioaugmentation: at a PCP concentration of 100 mg kg<sup>-1</sup>, the number of *C. testosteroni* CCM7530 CFUs rapidly increased over 17 days, while, at 10 mg PCP kg<sup>-1</sup>, the number of CFUs initially decreased until day 7 to then increase until day 17.

In any case, the level of selective pressure required to promote conjugal plasmid transfer depends on the specific contaminant and its concentration, as well as on the specific catabolic plasmid. In soil slurry, Ikuma and Gunsch (2012) observed that environmentally relevant concentrations of toluene might not exert enough pressure for transfer of plasmid TOL from *P. putida* BBC443 to *Serratia marcescens* and *P. fluorescens* cells. In their study on the degradation of 2,4-D, DiGiovanni et al. (1996) observed that this contaminant originated the required selective pressure for conjugal transfer of the intended catabolic plasmids.

Many biotic factors can also affect the success of plasmid-mediated bioaugmentation. Some genetic differences, such as guanine-cytosine (G+C) content and phylogenetic relationship between donor and recipient strain, can negatively affect the expression of the catabolic phenotype following conjugal plasmid transfer, as described by Ikuma and Gunsch (2012). Indeed, for plasmid-mediated bioaugmentation, biological differences between donor and recipient bacterial strains such as, for example, phylogenetic distance (Popa et al., 2011) and plasmid host range (De Gelder et al., 2005; Sorek et al., 2007), can play an important role. In 2,4-D contaminated soils, Newby et al. (2000) studied the bioaugmentation efficiency of two plasmid pJP4-bearing bacteria (the natural host, *Ralstonia eutropha* JMP134, and a laboratory-generated *E. coli* strain amenable to donor counterselection, named *E. coli* D11) and concluded that the correct choice of donor strain is a factor of the utmost importance for bioaugmentation.

Ikuma and Gunsch (2012) indicated that the success of plasmid-mediated bioaugmentation is dependent on: (i) high transfer rates of the catabolic plasmid to as many indigenous bacteria as possible; and (ii) the high expression level of an active contaminant-degrading phenotype in all transconjugants following conjugal plasmid transfer. Then, prior to the bioaugmentation process itself, it is important to characterize potentially recipient soil bacterial communities, paying special attention to dominant taxonomic groups. In the last years, next generation sequencing has provided a more comprehensive analysis of indigenous soil bacterial communities (Walsh, 2000), opening the door to the identification of potential recipient bacterial populations, and therefore a more informed selection of both the donor strain and the plasmid type (Ikuma and Gunsch, 2012).

Other biotic factors, such as competition between inoculated and indigenous bacteria for carbon sources, antagonistic interactions and predation by protozoa and bacteriophages, etc. also play an essential role in bioaugmentation efficiency. The critical factor is the selection of the right bacterial strains

(Thompson et al., 2005), since the inoculated strain must be able not only to degrade the target contaminant (or, in the case of plasmid-mediated bioaugmentation, to be able to effectively transfer the catabolic plasmid), but also to successfully compete with indigenous microbial populations and, in general, soil biota. On the other hand, plasmid transfer frequency has been shown to depend on the initial cell density ratio between donor and recipient cells (Pinedo and Smets, 2005; Ikuma et al., 2012).

Morphological, physiological and biochemical characteristics such as, for instance, cell size, growth rate, resource utilization ability, resistance phenotypes, biofilm formation capacity, cell motility, etc. are key traits for bacterial survival and competitiveness. Furthermore, DNA content has a marked influence on bacterial ecophysiological traits (i.e., adaptive traits to environmental changes) affecting, among other aspects, the rate of cell growth (Wickham and Lynn, 1990). Nevertheless, despite the assumption that fitness costs associated to HGT are caused by the need to maintain and replicate the extra-DNA, some studies indicate that they are predominantly due to transcription and translation processes (Bragg and Wagner, 2009; Shachrai et al., 2010).

The capacity of the host to use different carbon substrates before and after plasmid acquisition can provide an estimation of (i) its competitive ability and (ii) changes specifically associated to the plasmid transfer itself. Biolog<sup>TM</sup> plates can be employed to obtain a phenotypic fingerprint of bacterial strains in relation to their capacity to use a variety of carbon sources. Karve et al. (2016) followed phenotypic variations, using Biolog GEN III MicroPlates<sup>TM</sup>, to assess functionally relevant consequences of DNA changes.

Antibiotic resistance is probably the most extensively studied bacterial competitive trait. As a consequence of the production of antibiotics by soil microbial populations (D'Costa et al., 2006), soil is thought to be the largest reservoir of antibiotic resistance genes. Owing to fitness costs associated to antibiotic resistance, when bacteria change to an antibiotic-free environment, resistance is expected to disappear (Morosini et al., 2000), according to the assumption that, in the absence of selective pressure, resistant bacteria with a lower fitness will be outcompeted by susceptible counterparts with a higher fitness. However, it seems that bacteria tend to keep the mechanisms of antibiotic resistance, in order to maintain such an advantageous trait in the face of a possible change in environmental conditions (San Millan et al., 2014). Besides, in nature, antibiotics and antibiotic resistance determinants might play a variety of roles (e.g., signaling molecules in quorum sensing and biofilm formation, production of virulence factors, host-parasite interactions) (Sengupta et al., 2013) that justify the preservation of antibiotic resistance determinants in the absence of the selective pressure.

Biofilms are known to protect bacterial cells against antimicrobials (Høiby et al., 2010), predation, oxidative stress (Geier et al., 2008), etc. Biofilms harbor spatially structured bacterial communities where plasmids can be more easily shared through HGT (Jefferson, 2004), facilitating, for instance, the dissemination of catabolic genes. Remarkably, attachment to

surfaces by biofilm-associated factors is another cellular function associated to genes present in plasmids (Norman et al., 2009).

Cell motility is a critical aspect for the necessary dispersal of inoculated bacteria toward the target contaminants. Nevertheless, although highly motile bacterial cells, in their search for energy and nutrients, can disperse more easily into the surrounding environment, they also have a higher probability of encountering potential competitors (Reichenbach et al., 2007). In any case, motile bacterial populations, such as swarming bacteria, can more rapidly colonize new niches, with the associated ecological benefits (Verstraeten et al., 2008). Interestingly, there is a complex link between motility and biofilm formation because both processes appear to involve similar components at certain stages and conditions (Verstraeten et al., 2008).

(Gardin and Pauss, 2001; Gentili et al., 2006) have used different strategies of cell encapsulation and immobilization to facilitate inoculation survival, by providing a protective niche and temporary nutrition for the inoculated bacteria. Carrier materials, such as charcoal (Beck, 1991), nylon (Heitkamp and Steward, 1996), chitin, chitosan (Gentili et al., 2006; Chen et al., 2007) and zeolite (Liang et al., 2009) have been used in an attempt to maintain inoculant activity over a sufficiently long period of time after strain inoculation.

It must be taken into consideration that the influence of all these abovementioned abiotic and biotic factors has only been studied in a very limited number of bacterial strains and, in many cases, under controlled simplified environmental conditions, very different from those encountered in the natural environment. Therefore, many more in-depth studies on the impact of abiotic and biotic factors on cell and plasmid-mediated bioaugmentation are needed.

## CONCLUDING REMARKS

Both cell bioaugmentation and genetic (plasmid-mediated) bioaugmentation have proven effective for the bioremediation

of soils contaminated with organic compounds. However, cell bioaugmentation has an important limitation, i.e., the frequently very high mortality of the inoculated microbial strains, due to biotic or abiotic stresses. Then, *a priori*, plasmid-mediated bioaugmentation appears to have greater potential than cell bioaugmentation, since plasmids can act as gene-messenger biological tools that can transfer the required catabolic genes to indigenous bacterial populations already adapted to the soil under remediation. But for plasmid-mediated bioaugmentation to be successful and reproducible, much more research is needed for a better selection of donor bacterial strains and accompanying plasmids, together with an in-depth understanding of indigenous soil bacterial populations and the environmental conditions that affect plasmid acquisition and the expression and functioning of the catabolic genes of interest. Similarly, further research is required to better understand and then improve the ecological fitness of recipient bacterial strains in the contaminated soil.

## AUTHOR CONTRIBUTIONS

CG and IA: Design of the work and the acquisition of the data, writing and revision of the content, approval of the last version and ensuring accuracy and integrity of the work. LE and OG: Acquisition of the data, writing and revision of the content, approval of the last version of the work. EG: Writing and revision of the content, approval of the last version and ensuring accuracy and integrity of the work.

## ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Economy, Industry and Competitiveness (AGL2016-76592-R), and the Interreg SUDOE Programme (PhytoSUDOE-SOE1/P5/EO189). OG was a pre-doctoral student supported by the Basque Government and by the Fundación Biofísica Bizkaia.

## REFERENCES

- Aminov, R. I. (2011). Horizontal gene exchange in environmental microbiota. *Front. Microbiol.* 2:158. doi: 10.3389/fmicb.2011.0015
- Aspray, T. J., Hansen, S. K., and Burns, R. G. (2005). A soil-based microbial biofilm exposed to 2,4-D: bacterial community development and establishment of conjugative plasmid pJP4. *FEMS Microbiol. Ecol.* 54, 317–327. doi: 10.1016/j.femsec.2005.04.007
- Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* 28, 489–495. doi: 10.1016/j.tree.2013.04.002
- Barrios, E. (2007). Soil biota, ecosystem services and land productivity. *Ecol. Econ.* 64, 269–285. doi: 10.1016/j.ecolecon.2007.03.004
- Barrutia, O., Epelde, L., García-Plazaola, J. I., Garbisu, C., and Becerril, J. M. (2009). Phytoextraction potential of two *Rumex acetosa* L. accessions collected from metalliferous and non-metalliferous sites: effect of fertilization. *Chemosphere* 74, 259–264. doi: 10.1016/j.chemosphere.2008.09.036
- Barrutia, O., Garbisu, C., Epelde, L., Sampedro, M. C., Goicolea, M. A., and Becerril, J. M. (2011). Plant tolerance to diesel minimizes its impact on soil microbial characteristics during rhizoremediation of diesel-contaminated soils. *Sci. Total Environ.* 409, 4087–4093. doi: 10.1016/j.scitotenv.2011.06.025
- Barrutia, O., Garbisu, C., Hernández-Allica, J., García-Plazaola, J. I., and Becerril, J. M. (2010). Differences in EDTA-assisted metal phytoextraction between metalicolous and non-metallicolous accessions of *Rumex acetosa* L. *Environ. Pollut.* 158, 1710–1715. doi: 10.1016/j.envpol.2009.11.027
- Beck, D. P. (1991). Suitability of charcoal-amended mineral soil as carrier for *Rhizobium* inoculants. *Soil Biol. Biochem.* 23, 41–44. doi: 10.1016/0038-0717(91)90160-L
- Beil, S., Timmis, K. N., and Pieper, D. H. (1999). Genetic and biochemical analyses of the *tec* operon suggest a route for evolution of chlorobenzene degradation genes. *J. Bacteriol.* 181, 341–346.
- Bento, F. M., Camargo, F. A. O., Okeke, B. C., and Frankenberger, W. T. (2005). Comparative bioremediation of soils contaminated with diesel oil by natural attenuation, biostimulation and bioaugmentation. *Bioresour. Technol.* 96, 1049–1055. doi: 10.1016/j.biortech.2004.09.008
- Bertels, F., and Rainey, P. B. (2011). Within-genome evolution of REPINs: a new family of miniature mobile DNA in bacteria. *PLOS Genet.* 7:e1002132. doi: 10.1371/journal.pgen.1002132
- Besser, J. M., Brumbaugh, W. G., Allert, A. L., Poulton, B. C., and Schmitt, C. J. (2009). Ecological impacts of lead mining on Ozark streams-toxicity of sediment and pore water. *Ecotoxicol. Environ. Saf.* 72, 516–526. doi: 10.1016/j.ecoenv.2008.05.013
- Bouma, J. E., and Lenski, R. E. (1988). Evolution of a bacteria/plasmid association. *Nature* 335, 351–352. doi: 10.1038/335351a0

- Bragg, J. G., and Wagner, A. (2009). Protein material costs: single atoms can make an evolutionary difference. *Trends Genet.* 25, 5–8. doi: 10.1016/j.tig.2008.10.007
- Burges, A., Epelde, L., and Garbisu, C. (2015). Impact of repeated single-metal and multi-metal pollution events on soil quality. *Chemosphere* 120, 8–15. doi: 10.1016/j.chemosphere.2014.05.037
- Chen, Y. M., Lin, T. F., Huang, C., Lin, J. C., and Hsieh, F. M. (2007). Degradation of phenol and TCE using suspended and chitosan-bead immobilized *Pseudomonas putida*. *J. Hazard Mater.* 148, 660–670. doi: 10.1016/j.jhazmat.2007.03.030
- Cho, Y. G., Rhee, S. K., and Lee, S. T. (2000). Effect of soil moisture on bioremediation of chlorophenol-contaminated soil. *Biotechnol. Lett.* 22, 915–919. doi: 10.1023/A:1005612232079
- Chou, H. H., Chiu, H. C., Delaney, N. F., Segrè, D., and Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332, 1190–1192. doi: 10.1126/science.1203799
- Christensen, B. B., Sternberg, C., Andersen, J. B., Eberl, L., Møller, S., Givskov, M., et al. (1998). Establishment of new genetic traits in a microbial biofilm community. *Appl. Environ. Microbiol.* 64, 2247–2255.
- Davis, J. W., and Madsen, S. (1996). Factors affecting the biodegradation of toluene in soil. *Chemosphere* 33, 107–130. doi: 10.1016/0045-6535(96)00152-X
- D'Costa, V. M., McGrann, K. M., Hughes, D. W., and Wright, G. D. (2006). Sampling the antibiotic resistome. *Science* 311, 374–377. doi: 10.1126/science.1120800
- De Gelder, L., Ponciano, J. M., Joyce, P., and Top, E. M. (2007). Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology* 153, 452–463. doi: 10.1099/mic.0.2006/001784-0
- De Gelder, L., Vandecasteele, F. P. J., Brown, C. J., Forney, L. J., and Top, E. M. (2005). Plasmid donor affects host range of promiscuous IncP-1 $\beta$  plasmid pB10 in an activated-sludge microbial community. *Appl. Environ. Microbiol.* 71, 5309–5317. doi: 10.1128/AEM.71.9.5309-5317.2005
- Dejonghe, W., Goris, J., El Fantroussi, S., Hofte, M., De Vos, P., Verstraete, W., et al. (2000). Effect of dissemination of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation plasmids on 2,4-D degradation and on bacterial community structure in two different soil horizons. *Appl. Environ. Microbiol.* 66, 3297–3304. doi: 10.1128/AEM.66.8.3297-3304.2000
- DiGiovanni, G. D., Neilson, J. W., Pepper, I. L., and Sinclair, N. A. (1996). Gene transfer of *Alcaligenes eutrophus* JMP134 plasmid pJP4 to indigenous soil recipients. *Appl. Environ. Microbiol.* 62, 2521–2526.
- Dionisio, F., Conceição, I. C., Marques, A. C., Fernandes, L., and Gordo, I. (2005). The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biol. Lett.* 1, 250–252. doi: 10.1098/rsbl.2004.0275
- Directive 2001/18/EC (2001). *Directive/18/EC of the European Parliament and of the Council of 12 March 2001 on the Deliberate Release into the Environment of Genetically Modified Organisms and Repealing Council Directive 90/220/EEC*. Brussels: European Commission.
- Djokic, L., Narancic, T., Biocanic, M., Saljnikov, E., Casey, E., Vasiljevic, B., et al. (2013). Phenol removal from four different natural soil types by *Bacillus* sp. PS11. *Appl. Soil Ecol.* 70, 1–8. doi: 10.1016/j.apsoil.2013.04.002
- El Fantroussi, S., and Agathos, S. N. (2005). Is bioaugmentation a feasible strategy for pollutant removal and site remediation? *Curr. Opin. Microbiol.* 8, 268–275. doi: 10.1016/j.mib.2005.04.011
- Epelde, L., Becerril, J. M., Alkorta, I., and Garbisu, C. (2009). “Heavy metal phytoremediation: microbial indicators of soil health for the assessment of remediation efficiency,” in *The Advances in Applied Bioremediation*, eds A. Singh, R. C. Kuhad, and O. P. Ward (Berlin: Springer), 299–313.
- Epelde, L., Becerril, J. M., Alkorta, I., and Garbisu, C. (2014). Adaptive long-term monitoring of soil health in metal phytostabilization: ecological attributes and ecosystem services based on soil microbial parameters. *Int. J. Phytoremediat.* 16, 971–981. doi: 10.1080/15226514.2013.810578
- Epelde, L., Becerril, J. M., Kowalchuk, G. A., Deng, Y., Zhou, J., and Garbisu, C. (2010). Impact of metal pollution and *Thlaspi caerulescens* growth on soil microbial communities. *Appl. Environ. Microbiol.* 76, 7843–7853. doi: 10.1128/AEM.01045-10
- Filonov, A. E., Akhmetov, L. I., Puntus, I. F., Esikova, T. Z., Gafarov, A. B., Kosheleva, I. A., et al. (2010). Horizontal transfer of catabolic plasmids and naphthalene biodegradation in open soil. *Microbiology* 79, 184–190. doi: 10.1134/S0026261710020098
- Franzetti, A., Caredda, P., Ruggeri, C., La Colla, L., Tamburini, E., Papacchini, M., et al. (2009). Potential application of surface active compounds by *Gordonia* sp. strain BS29 in soil remediation technologies. *Chemosphere* 75, 801–807. doi: 10.1016/j.chemosphere.2008.12.052
- Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev.* 3, 722–732. doi: 10.1038/nrmicro1235
- Furuya, E. Y., and Lowy, F. D. (2006). Antimicrobial-resistant bacteria in the community setting. *Nat. Rev. Microbiol.* 4, 36–45. doi: 10.1038/nrmicro1325
- Gao, C., Jin, X., Ren, J., Fang, H., and Yu, Y. (2015). Bioaugmentation of DDT-contaminated soil by dissemination of the catabolic plasmid pDOD. *J. Environ. Sci.* 27, 42–50. doi: 10.1016/j.jes.2014.05.045
- Garbisu, C., and Alkorta, I. (1997). Utilization of genetically engineered microorganisms (GEMs) for bioremediation. *J. Chem. Technol. Biotechnol.* 74, 599–606. doi: 10.1080/07388550600842794
- Garbisu, C., Alkorta, I., and Epelde, L. (2011). Assessment of soil quality using microbial properties and attributes of ecological relevance. *Appl. Soil Ecol.* 49, 1–4. doi: 10.1016/j.apsoil.2011.04.018
- Garbisu, C., Hernández-Allica, J., Barrutia, O., Alkorta, I., and Becerril, J. M. (2002). Phytoremediation: a technology using green plants to remove contaminants from polluted areas. *Rev. Environ. Health* 17, 173–188. doi: 10.1515/REVEH.2002.17.3.173
- Gardin, H., and Paus, A. (2001).  $\kappa$ -carrageenan/gelatin gel beads for the co-immobilization of aerobic and anaerobic microbial communities degrading 2,4,6-trichlorophenol under air-limited conditions. *Appl. Microbiol. Biotechnol.* 56, 517–523. doi: 10.1007/s002530000581
- Geier, H., Mostowy, S., Cangelosi, G. A., Behr, M. A., and Ford, T. E. (2008). Autoinducer-2 triggers the oxidative stress response in *Mycobacterium avium*, leading to biofilm formation. *Appl. Environ. Microbiol.* 74, 1798–1804. doi: 10.1128/AEM.02066-07
- Gentili, A. R., Cubitto, M. A., Ferrero, M., and Rodríguez, M. S. (2006). Bioremediation of crude oil polluted seawater by a hydrocarbon-degrading bacterial strain immobilized on chitin and chitosan flakes. *Int. Biodeterior. Biodegrad.* 57, 222–228. doi: 10.1016/j.ibiod.2006.02.009
- Gómez-Sagasti, M. T., Epelde, L., Alkorta, I., and Garbisu, C. (2016). Reflections on soil contamination research from a biologist's point of view. *Appl. Soil Ecol.* 105, 207–210. doi: 10.1016/j.apsoil.2016.04.004
- Greer, L. E., and Shelton, D. R. (1992). Effect of inoculant strain and organic matter content on kinetics of 2,4-dichloro-phenoxyacetic acid degradation in soil. *Appl. Environ. Microbiol.* 58, 1459–1465.
- Halden, R. U., Tepp, S., Halden, B. G., and Dwyer, D. F. (1999). Degradation of 3-phenoxybenzoic acid in soil by *Pseudomonas pseudoalcaligenes* POB310(pPOB) and two modified *Pseudomonas* strains. *Appl. Environ. Microbiol.* 65, 3354–3359.
- Hall, J. P. J., Brockhurst, M. A., Dytham, C., and Harrison, E. (2017). The evolution of plasmid stability: are infectious transmission and compensatory evolution competing evolutionary trajectories. *Plasmid* 91, 90–95. doi: 10.1016/j.plasmid.2017.04.003
- Hall, J. P. J., Wood, A. J., Harrison, E., and Brockhurst, M. A. (2016). Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 113, 8260–8265. doi: 10.1073/pnas.1600974113
- Haluška, L., Barančíková, G., Baláž, Š., Dercová, K., Vrana, B., Paz-Weisshaar, M., et al. (1995). Degradation of PCB in different soils by inoculated *Alcaligenes xyloxydans*. *Sci. Total Environ.* 175, 275–285. doi: 10.1016/0048-9697(95)04927-4
- Hansch, C., McKarns, S. C., Smith, C. J., and Doolittle, D. J. (2000). Comparative QSAR evidence for a free-radical mechanism of phenol-induced toxicity. *Chem. Biol. Interact.* 127, 61–72. doi: 10.1016/S0009-2797(00)00171-x
- Harrison, E., Guymer, D., Spiers, A. J., Paterson, S., and Brockhurst, M. A. (2015). Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr. Biol.* 25, 2034–2039. doi: 10.1016/j.cub.2015.06.024
- Heinrich, E., Merimaa, M., Viggior, S., Lehist, M., Leito, I., Truu, J., et al. (2005). Biodegradation efficiency of functionally important populations selected for bioaugmentation in phenol and oil-polluted area. *FEMS Microbiol. Ecol.* 51, 363–373. doi: 10.1016/j.femsec.2004.09.009



- Heitkamp, M. A., and Steward, W. P. (1996). A novel porous nylon biocarrier for immobilized bacteria. *Appl. Environ. Microbiol.* 62, 4659–4662.
- Høiby, N., Bjarnsholt, T., Givskov, M., Molin, S., and Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *Int. J. Antimicrob. Agents* 35, 322–332. doi: 10.1016/j.ijantimicag.2009.12.011
- Ikuma, K., and Gunsch, C. K. (2010). Effect of carbon source addition on toluene biodegradation by an *Escherichia coli* DH5 $\alpha$  transconjugant harboring the TOL plasmid. *Biotechnol. Bioeng.* 107, 269–277. doi: 10.1002/bit.22808
- Ikuma, K., and Gunsch, C. K. (2012). Genetic bioaugmentation as an effective method for *in situ* bioremediation: functionality of catabolic plasmids following conjugal transfers. *Bioengineered* 3, 236–241. doi: 10.4161/bioe.20551
- Ikuma, K., Holzem, R. M., and Gunsch, C. K. (2012). Impacts of organic carbon availability and recipient bacteria characteristics on the potential for TOL plasmid genetic bioaugmentation in soil slurries. *Chemosphere* 89, 158–163. doi: 10.1016/j.chemosphere.2012.05.086
- Inoue, D., Sei, K., Soda, S., Ike, M., and Fujita, M. (2005). Potential of predominant activated sludge bacteria as recipients in conjugative plasmid transfer. *J. Biosci. Bioeng.* 100, 600–605. doi: 10.1263/jbb.100.600
- Inoue, D., Yamazaki, Y., Tsutsui, T., Sei, K., Soda, S., Fujita, M., et al. (2012). Impacts of gene bioaugmentation with pJP4-harboring bacteria of 2,4-D-contaminated soil slurry on the indigenous microbial community. *Biodegradation* 23, 263–276. doi: 10.1007/s10532-011-9505-x
- Iwamoto, T., and Nasu, M. (2001). Current bioremediation practice and perspective. *J. Biosci. Bioeng.* 92, 1–8. doi: 10.1016/S1389-1723(01)80190-0
- Jefferson, K. K. (2004). What drives bacteria to produce a biofilm? *FEMS Microbiol. Lett.* 236, 163–173. doi: 10.1016/j.femsle.2004.06.005
- Johnsen, A. R., and Kroer, N. (2007). Effects of stress and other environmental factors on horizontal plasmid transfer assessed by direct quantification of discrete transfer events. *FEMS Microbiol. Ecol.* 59, 718–728. doi: 10.1111/j.1574-6941.2006.00230.x
- Jung, C. M., Crocker, F. H., Eberly, J. O., and Indest, K. J. (2011). Horizontal gene transfer (HGT) as a mechanism of disseminating RDX-degrading activity among *Actinomyces* bacteria. *J. Appl. Microbiol.* 110, 1449–1459. doi: 10.1111/j.1365-2672.2011.04995.x
- Jussila, M. M., Zhao, J., Suominen, L., and Lindström, K. (2007). TOL plasmid transfer during bacterial conjugation *in vitro* and rhizoremediation of oil compounds *in vivo*. *Environ. Pollut.* 146, 510–524. doi: 10.1016/j.envpol.2006.07.012
- Juwarkar, A. A., Misra, R. R., and Sharma, J. K. (2014). “Recent trends in bioremediation,” in *Geomicrobiology and Biogeochemistry*, eds N. Parmar and A. Singh (Berlin: Springer), 81–100.
- Karve, S. M., Tiwary, K., Selveshwari, S., and Dey, S. (2016). Environmental fluctuations do not select for increased variation or population-based resistance in *Escherichia coli*. *J. Biosci.* 41, 39–49. doi: 10.1007/s12038-016-9592-2
- Kim, J. M., Le, N. T., Chung, B. S., Park, J. H., Bae, J. W., Madsen, E. L., et al. (2008). Influence of soil components on the biodegradation of benzene, toluene, ethylbenzene, and *o*-, *m*-, and *p*-xylene by the newly isolated bacterium *Pseudoxanthomonas spadix* BD-a59. *Appl. Environ. Microbiol.* 74, 7313–7320. doi: 10.1128/AEM.01695-08
- Kottara, A., Hall, J. P. J., Harrison, E., and Brockhurst, M. A. (2016). Multi-host environments select for host-generalist conjugative plasmids. *BMC Evol. Biol.* 16:70. doi: 10.1186/s12862-016-0642-z
- Król, J. E., Nguyen, H. D., Rogers, L. M., Beyenal, H., Krone, S. M., and Top, E. M. (2011). Increased transfer of a multidrug resistance plasmid in *Escherichia coli* biofilms at the air-liquid interface. *Appl. Environ. Microbiol.* 77, 5079–5088. doi: 10.1128/AEM.00090-11
- Kumavath, R. N., and Deverapalli, P. (2013). “Scientific swift in bioremediation: an overview,” in *Applied Bioremediation - Active and Passive Approaches*, eds Y. B. Patil and P. Rao (Croatia: Intech), 375–388.
- Liang, Y., Zhang, X., Dai, D., and Li, G. (2009). Porous biocarrier-enhanced biodegradation of crude oil contaminated soil. *Int. Biodeterior. Biodegrad.* 63, 80–87. doi: 10.1016/j.ibiod.2008.07.005
- Maier, R. M. (2000). “Microorganisms and organic pollutants,” in *Environmental Microbiology*, eds R. M. Maier, I. L. Pepper, and C. P. Gerba (San Diego, CA: Academic Press), 63–402.
- Maphosa, F., Lieten, S. H., Dinkla, I., Stams, A. J., Smidt, H., and Fennell, D. E. (2012). Ecogenomics of microbial communities in bioremediation of chlorinated contaminated sites. *Front. Microbiol.* 3:351. doi: 10.3389/fmicb.2012.0035
- Martínez, J. L., Sánchez, M. B., Martínez-Solano, L., Hernández, A., Garmendia, L., Fajardo, A., et al. (2009). Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol. Rev.* 33, 430–449. doi: 10.1111/j.1574-6976.2008.00157.x
- Massa, V., Infantino, A., Radice, F., Orlandi, V., Tavecchio, F., Giudici, R., et al. (2009). Efficiency of natural and engineered bacterial strains in the degradation of 4-chlorobenzoic acid in soil slurry. *Int. Biodeterior. Biodegrad.* 63, 112–125. doi: 10.1016/j.ibiod.2008.07.006
- Meckenstock, R. U., Elsner, M., Griebler, C., Lueders, T., Stumpp, C., Aamand, J., et al. (2015). Biodegradation: updating the concepts of control for microbial cleanup in contaminated aquifers. *Environ. Sci. Technol.* 49, 7073–7081. doi: 10.1021/acs.est.5b00715
- Miller, M. N., Stratton, G. W., and Murray, G. (2004). Effects of soil moisture and aeration on the biodegradation of pentachlorophenol contaminated soil. *Bull. Environ. Contam. Toxicol.* 72, 101–108. doi: 10.1007/s00128-003-0246-3
- Miyazaki, R., Sato, Y., Ito, M., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2006). Complete nucleotide sequence of an exogenously isolated plasmid, pLB1, involved in gamma-hexachlorocyclohexane degradation. *Appl. Environ. Microbiol.* 72, 6923–6933. doi: 10.1128/AEM.01531-06
- Mølbak, L., Molin, S., and Kroer, N. (2007). Root growth and exudate production define the frequency of horizontal plasmid transfer in the rhizosphere. *FEMS Microbiol. Ecol.* 59, 167–176. doi: 10.1111/j.1574-6941.2006.00229.x
- Morosini, M. I., Ayala, J. A., Baquero, F., Martínez, J. L., and Blázquez, J. (2000). Biological cost of AmpC production for *Salmonella enterica* serotype Typhimurium. *Antimicrob. Agents Chemother.* 44, 3137–3143. doi: 10.1128/AAC.44.11.3137-3143.2000
- Mrozik, A., Miga, S., and Piotrowska-Seget, Z. (2011). Enhancement of phenol degradation by soil bioaugmentation with *Pseudomonas* sp. JS150. *J. Appl. Microbiol.* 111, 1357–1370. doi: 10.1111/j.1365-2672.2011.05140.x
- Mrozik, A., and Piotrowska-Seget, Z. (2010). Bioaugmentation as a strategy for cleaning up of soils contaminated with aromatic compounds. *Microbiol. Res.* 165, 363–375. doi: 10.1016/j.micres.2009.08.001
- Muñoz-Leoz, B., Garbisu, C., Charcosset, J. Y., Sánchez-Pérez, J. M., Antigüedad, I., Ruiz, E., et al. (2013). Non-target effects of three formulated pesticides on microbially-mediated processes in a clay-loam soil. *Sci. Total Environ.* 449, 345–354. doi: 10.1016/j.scitotenv.2013.01.079
- Musovic, S., Dechesne, A., Sørensen, J., and Smets, B. F. (2010). A novel assay to assess permissiveness of a soil microbial community toward receipt of mobile genetic elements. *Appl. Environ. Microbiol.* 76, 4813–4818. doi: 10.1128/AEM.02713-09
- Newby, D. T., Gentry, T. J., and Pepper, I. L. (2000). Comparison of 2,4-dichlorophenoxyacetic acid degradation and plasmid transfer in soil resulting from bioaugmentation with two different pJP4 donors. *Appl. Environ. Microbiol.* 66, 3399–3407. doi: 10.1128/AEM.66.8.3399-3407.2000
- Newby, D. T., and Pepper, I. L. (2002). Dispersal of plasmid pJP4 in unsaturated and saturated 2,4-dichlorophenoxyacetic acid contaminated soil. *FEMS Microbiol. Ecol.* 39, 157–164. doi: 10.1111/j.1574-6941.2002.tb00917.x
- Norman, A., Hansen, L. H., and Sørensen, S. J. (2009). Conjugative plasmids: vessels of the communal gene pool. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 364, 2275–2289. doi: 10.1098/rstb.2009.0037
- Nowak, A., and Mrozik, A. (2016). Facilitation of co-metabolic transformation and degradation of monochlorophenols by *Pseudomonas* sp. CF600 and changes in its fatty acid composition. *Water Air Soil Pollut.* 227, 83. doi: 10.1007/s11270-016-2775-5
- Overhage, J., Selker, S., Homburg, S., Parschat, K., and Fetzner, S. (2005). Identification of large linear plasmids in *Arthrobacter* spp. encoding the degradation of quinaldine to anthranilate. *Microbiology* 151, 491–500. doi: 10.1099/mic.0.27521-0
- Panagos, P., Van Liedekerke, M., Yigini, Y., and Montanarella, L. (2013). Contaminated sites in Europe: review of the current situation based on data collected through a European network. *J. Environ. Public Health* 2013, 1–11. doi: 10.1155/2013/158764
- Pardo, T., Clemente, R., Epelde, L., Garbisu, C., and Bernal, M. P. (2014). Evaluation of the phytostabilisation efficiency in a trace elements contaminated soil using soil health indicators. *J. Hazard Mater.* 268, 68–76. doi: 10.1016/j.jhazmat.2014.01.003



- Park, C., and Zhang, J. (2012). High expression hampers horizontal gene transfer. *Genome Biol. Evol.* 4, 523–532. doi: 10.1093/gbe/evs030
- Pepper, I. L., Gentry, T. J., Newby, D. T., Roane, T. M., and Josephson, K. L. (2002). The role of cell bioaugmentation and gene bioaugmentation in the remediation of co-contaminated soils. *Environ. Health Perspect.* 110, 943–946. doi: 10.1289/ehp.02110s6943
- Pinedo, C. A., and Smets, B. F. (2005). Conjugal TOL transfer from *Pseudomonas putida* to *Pseudomonas aeruginosa*: effects of restriction proficiency, toxicant exposure, cell density ratios, and conjugation detection method on observed transfer efficiencies. *Appl. Environ. Microbiol.* 71, 51–57. doi: 10.1128/AEM.71.1.51-57.2005
- Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609. doi: 10.1101/gr.115592.110
- Popowska, M., and Krawczyk-Balska, A. (2013). Broad-host-range IncP-1 plasmids and their resistance potential. *Front. Microbiol.* 4:44. doi: 10.3389/fmicb.2013.00044
- Porse, A., Schønning, K., Munck, C., and Sommer, M. O. A. (2016). Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. *Mol. Biol. Evol.* 33, 2860–2873. doi: 10.1093/molbev/msw163
- Quan, X., Tang, H., Xiong, W., and Yang, Z. (2010). Bioaugmentation of aerobic sludge granules with a plasmid donor strain for enhanced degradation of 2,4-dichlorophenoxyacetic acid. *J. Hazard Mater.* 179, 1136–1142. doi: 10.1016/j.jhazmat.2010.04.00
- Reichenbach, T., Mobilia, M., and Frey, E. (2007). Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. *Nature* 448, 1046–1049. doi: 10.1038/nature06095
- Reineke, W. (1998). Development of hybrid strains for the mineralization of chloroaromatics by patchwork assemble. *Annu. Rev. Microbiol.* 52, 287–331. doi: 10.1146/annurev.micro.52.1.287
- Riley, M. A., and Wertz, J. E. (2002). Bacteriocins: evolution, ecology and application. *Annu. Rev. Microbiol.* 56, 117–137. doi: 10.1146/annurev.micro.56.012302.161024
- San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z. V., McLean, A. R., Cooper, B. S., et al. (2014). Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nat. Commun.* 5, 5208. doi: 10.1038/ncomms6208
- Sayler, G. S., and Ripp, S. (2000). Field applications of genetically engineered microorganisms for bioremediation processes. *Curr. Opin. Biotechnol.* 11, 286–289. doi: 10.1016/S0958-1669(00)00097-5
- Sejáková, Z., Dercová, K., and Tóthová, L. (2009). Biodegradation and ecotoxicity of soil contaminated by pentachlorophenol applying bioaugmentation and addition of sorbents. *World J. Microbiol. Biotechnol.* 25, 243–252. doi: 10.1007/s11274-008-9885-1
- Sengupta, S., Chattopadhyay, M. K., and Grossart, H. P. (2013). The multifaceted roles of antibiotics and antibiotic resistance in nature. *Front. Microbiol.* 4:47. doi: 10.3389/fmicb.2013.00047
- Shachrai, I., Zaslaver, A., Alon, U., and Dekel, E. (2010). Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Mol. Cell* 38, 758–767. doi: 10.1016/j.molcel.2010.04.015
- Shintani, M., Sanchez, Z. K., and Kimbara, K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front. Microbiol.* 31:242. doi: 10.3389/fmicb.2015.00242
- Shintani, M., Takahashi, Y., Yamane, H., and Nojiri, H. (2010a). The behavior and significance of degradative plasmids belonging to Inc groups in *Pseudomonas* within natural environments and microcosms. *Microbes Environ.* 25, 253–265. doi: 10.1264/jsme2.ME10155
- Shintani, M., Yamane, H., and Nojiri, H. (2010b). Behaviour of various hosts of the IncP-7 carbazole-degradative plasmid pCAR1 in artificial microcosms. *Biosci. Biotechnol. Biochem.* 74, 343–349. doi: 10.1271/bbb.90675
- Singh, A., and Ward, O. P. (2004). “Biotechnology and bioremediation An overview,” in *Biodegradation and Bioremediation*, eds A. Singh and O. P. Ward (Berlin: Springer-Verlag), 1–17.
- Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., and Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449–1452. doi: 10.1126/science.1147112
- Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N., and Wuertz, S. (2005). Studying plasmid horizontal transfer *in situ*: a critical review. *Nat. Rev. Microbiol.* 3, 700–710. doi: 10.1038/nrmicro1232
- Stalwood, B., Shears, J., Williams, P. A., and Hughes, K. A. (2005). Low temperature bioremediation of oil-contaminated soil using biostimulation and bioaugmentation with a *Pseudomonas* sp. from maritime Antarctica. *J. Appl. Microbiol.* 99, 794–802. doi: 10.1111/j.1365-2672.2005.02678.x
- Starikova, I., Al-Haroni, M., Werner, G., Roberts, A., Sørum, V., Nielsen, K. M., et al. (2013). Fitness costs of various mobile genetic elements in *Enterococcus faecium* and *Enterococcus faecalis*. *J. Antimicrob. Chemother.* 68, 2755–2765. doi: 10.1093/jac/dkt270
- Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi: 10.1038/nrmicro1234
- Thompson, I. P., Van der Gast, C. J., Ciric, L., and Singer, A. C. (2005). Bioaugmentation for bioremediation: the challenge of strain selection. *Environ. Microbiol.* 7, 909–915. doi: 10.1111/j.1462-2920.2005.00804.x
- Top, E. M., and Springael, D. (2003). The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr. Opin. Biotechnol.* 14, 262–269. doi: 10.1016/S0958-1669(03)00066-1
- Top, E. M., Springael, D., and Boon, N. (2002). Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiol. Ecol.* 42, 199–208. doi: 10.1111/j.1574-6941.2002.tb01009.x
- Top, E. M., Van Elsas, E. D., DeSmet, S. N., and Forney, L. J. (1998). Enhancement of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation in soil by dissemination of catabolic plasmids. *Antonie Van Leeuwenhoek* 73, 87–94. doi: 10.1023/A:1000663619522
- Tsutsui, H., Anami, Y., Matsuda, M., Hashimoto, K., Inoue, D., Sei, K., et al. (2013). Plasmid-mediated bioaugmentation of sequencing batch reactors for enhancement of 2,4-dichlorophenoxyacetic acid removal in wastewater using plasmid pJP4. *Biodegradation* 24, 343–352. doi: 10.1007/s10532-012-9591-4
- van Rensburg, E., den Haan, R., Smith, J., van Zyl, W. H., and Görgens, J. F. (2012). The metabolic burden of cellulase expression by recombinant *Saccharomyces cerevisiae* Y294 in aerobic batch culture. *Appl. Microbiol. Biotechnol.* 96, 197–209. doi: 10.1007/s00253-012-4037-9
- Venkata Mohan, S., Falkentoft, C., Nancharaiya, Y. V., Sturm, B. S. M., Wattiau, P., Wilderer, P. A., et al. (2009). Bioaugmentation of microbial communities in laboratory and pilot scale sequencing batch biofilm reactors using the TOL plasmid. *Bioresour. Technol.* 100, 1746–1753. doi: 10.1016/j.biortech.2008.09.048
- Verstraeten, N., Braeken, K., Debkumari, B., Fauvart, M., Franssaer, J., Vermant, J., et al. (2008). Living on a surface: swarming and biofilm formation. *Trends Microbiol.* 16, 496–506. doi: 10.1016/j.tim.2008.07.004
- Vogwill, T., and MacLean, R. C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol. Appl.* 8, 284–295. doi: 10.1111/eva.12202
- Walsh, C. (2000). Molecular mechanisms that confer antibacterial drug resistance. *Nature* 406, 775–781. doi: 10.1038/35021219
- Wang, J. L., Mao, Z. Y., Han, L. P., and Qian, Y. (2004). Bioremediation of quinoline-contaminated soil using bioaugmentation in slurry-phase reactor. *Biomed. Environ. Sci.* 17, 187–195.
- Wang, S., and Mulligan, C. N. (2004). An evaluation of surfactant foam technology in remediation of contaminated soil. *Chemosphere* 57, 1079–1089. doi: 10.1016/j.chemosphere.2004.08.019
- Wang, Y., Kou, S., Jiang, Q., Xu, B., Liu, X., Xiao, J., et al. (2014). Factors affecting transfer of degradative plasmids between bacteria in soils. *Appl. Soil Ecol.* 84, 254–261. doi: 10.1016/j.apsoil.2014.07.009
- Wickham, S. A., and Lynn, D. H. (1990). Relation between growth rate, cell size, and DNA content in colpodean ciliates (Ciliophora: Colpodea). *Eur. J. Protistol.* 25, 345–352. doi: 10.1016/S0932-4739(11)80127-6
- Wiedenbeck, J., and Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* 35, 957–976. doi: 10.1111/j.1574-6976.2011.00292.x

- Wolski, E. A., Murialdo, S. E., and Gonzales, J. F. (2006). Effect of pH and inoculum size on pentachlorophenol degradation by *Pseudomonas* sp. *Water SA* 32, 93–98.
- Wozniak, R. A., and Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8, 552–563. doi: 10.1038/nrmicro2382
- Wyndham, R. C., Cashore, A. E., Nakatsu, C. H., and Peel, M. C. (1994). Catabolic transposons. *Biodegradation* 5, 323–357. doi: 10.1007/BF00696468
- Zhang, Q., Wang, B. C., Cao, Z. Y., and Yu, Y. L. (2012). Plasmid-mediated bioaugmentation for the degradation of chlorpyrifos in soil. *J. Hazard. Mater.* 221–222, 178–184. doi: 10.1016/j.jhazmat.2012.04.024

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Garbisu, Garaiyurrebaso, Epelde, Grohmann and Alkorta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparative Genomic Analysis Reveals Organization, Function and Evolution of *ars* Genes in *Pantoea* spp.

Liying Wang<sup>1,2</sup>, Jin Wang<sup>3</sup> and Chuanyong Jing<sup>1,2\*</sup>

<sup>1</sup> State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China, <sup>2</sup> College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, <sup>3</sup> Department of Municipal and Environmental Engineering, School of Civil Engineering, Beijing Jiaotong University, Beijing, China

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### Reviewed by:

Yunyoung Kwak,  
Kyungpook National University,  
South Korea  
Lukasz Drewniak,  
University of Warsaw, Poland  
Ji-Hoon Lee,  
Chonbuk National University,  
South Korea

### \*Correspondence:

Chuanyong Jing  
cyjing@rcees.ac.cn

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 04 November 2016

**Accepted:** 07 March 2017

**Published:** 21 March 2017

### Citation:

Wang L, Wang J and Jing C (2017)  
Comparative Genomic Analysis  
Reveals Organization, Function  
and Evolution of *ars* Genes in *Pantoea*  
spp. *Front. Microbiol.* 8:471.  
doi: 10.3389/fmicb.2017.00471

Numerous genes are involved in various strategies to resist toxic arsenic (As). However, the As resistance strategy in genus *Pantoea* is poorly understood. In this study, a comparative genome analysis of 23 *Pantoea* genomes was conducted. Two vertical genetic *arsC*-like genes without any contribution to As resistance were found to exist in the 23 *Pantoea* strains. Besides the two *arsC*-like genes, As resistance gene clusters *arsRBC* or *arsRBCH* were found in 15 *Pantoea* genomes. These *ars* clusters were found to be acquired by horizontal gene transfer (HGT) from sources related to *Franconibacter helveticus*, *Serratia marcescens*, and *Citrobacter freundii*. During the history of evolution, the *ars* clusters were acquired more than once in some species, and were lost in some strains, producing strains without As resistance capability. This study revealed the organization, distribution and the complex evolutionary history of As resistance genes in *Pantoea* spp.. The insights gained in this study improved our understanding on the As resistance strategy of *Pantoea* spp. and its roles in the biogeochemical cycling of As.

**Keywords:** comparative genomic, arsenic, *Pantoea* spp., arsenic resistance, *ars* genes

## INTRODUCTION

Arsenic (As), one of the earliest known toxic elements, occurs naturally worldwide (Smith et al., 2002). To adapt to habitats with elevated As, microbes have evolved dynamic resistance mechanisms. The most ubiquitous and important strategy of As resistance is to reduce As(V) to As(III) and extrude it using *ars* operons with various genomic configurations in specific bacterial strains (Páezespino et al., 2009). The core genes of *ars* systems, however, are *arsR* (encoding the transcriptional repressor ArsR), *arsB* (encoding the arsenite efflux pump ArsB) and *arsC* (encoding arsenate reductase ArsC) (Xu et al., 1998). Besides this detoxification mechanism using *ars* systems, some strains possess the mechanism of As methylation-demethylation, changing inorganic As into organic forms using a distinct gene *arsM* (Qin et al., 2006; Zhao et al., 2015). Some strains are able to oxidize As(III) to As(V), which involve membrane-associated proteins, AoxAB (Levin and Tal, 2003; Ghosh et al., 2014). Some strains are able to reduce As(V) to As(III) with ArrAB as part of their respiratory processes transferring electrons to As and producing the energy for strains (Saltikov and Newman, 2003). The reported genes related to the strategy of As resistance are listed in Table 1.

In traditional molecular biology research, As resistance traits are revealed primarily based on the cultivation of a specific strain, and it is impossible to study the As strategy of all strains in a genus. Nevertheless, understanding of such traits in all strains of a genus is sometime more desirable. Gaining this knowledge is no longer a challenge with the explosive development of high-throughput sequencing technology. The genomic sequence of a strain contains nearly all of the genetic information. Therefore, fundamental knowledge such as the phylogenetic, the genetic traits of As resistance and its evolutionary history can be obtained through comparative genomic analysis (Arsène-Ploetze et al., 2010; Colston et al., 2014). So here we use genomic information of *Pantoea* spp. and compared these genomes to explore and predict the strategy of As resistance and their evolutionary patterns in genus *Pantoea* as an example.

*Pantoea* is a genus of Gram-negative, facultative anaerobic bacteria. This genus belongs to gamma *Proteobacteria*, family *Gammaproteobacteria*, and was recently separated from the genus *Enterobacter* (Gavini et al., 1989). Currently, the genus contains 26 species<sup>1</sup>. Members of this genus are found in various environmental matrices (Meng et al., 1995; Zhang and Birch, 1997; Rezzonico et al., 2009). In 2013, the strain *Pantoea* sp. IMH was an isolate that reported firstly as the strain having the As resistance capability within *Pantoea* species (Wu et al., 2013). Further, we sequenced the genome of *Pantoea* sp. IMH and found two *ars* clusters (*arsR1B1C1H1* and *arsR2B2C2H2*) co-contributing to its As resistance (Tian and Jing, 2014; Wang et al., 2016). However, the evolutionary history and genetic traits of As resistance in genus *Pantoea* are not fully understood.

Herein, we present the first study of the genetic traits of As resistance in *Pantoea* spp., as well as their evolutionary

history. Two vertically transmitted *arsC*-like genes without any contribution to As resistance were found to exist in the 23 *Pantoea* strains. Besides these two *arsC*-like genes, As resistance gene clusters *arsRBC* or *arsRBCH* were found in 15 *Pantoea* genomes. These *ars* clusters were acquired by horizontal gene transfer (HGT) from sources related to *Franconibacter helveticus*, *Serratia marcescens*, and *Citrobacter freundii*. The insights gained in this study improve our understanding on the complex evolutionary history of As resistance genes and their roles in the biogeochemical cycling of As.

## MATERIALS AND METHODS

### Phylogenetic Analysis

Phylogenetic trees of *Pantoea* species were constructed based on 100 single-copy core proteins shared by 23 *Pantoea* genomes and the genome of *Tatumella* sp. NML 06-3099 according to the following three methods: maximum likelihood (ML), neighbor joining (NJ), and Bayesian inference (BI). ML and NJ trees were computed by applying models with 1,000 bootstrap replicates and uniform rates in MEGA5 (Tamura, 2011). Multiple alignments of amino acid sequences were carried out by ClustalW, and the CONSEL program was used to select the best model of the trees (Shimodaira and Hasegawa, 2001; Thompson et al., 2002). The BI tree was generated using the MrBayes package with mixed models (Ronquist et al., 2012). The NJ tree of concatenated *arsRBC* homologs was generated according to the same method described above. MEGA5 or FigTree v.1.3.1<sup>2</sup> was used to illustrate the constructed trees.

### Average Nucleotide Identity (ANI)

Assembled contigs were reconstituted from the RAST-generated GenBank files for 23 genomes by using the seqret function of the EMBOSS package (Rice et al., 2000). These 23 genomes were treated in the same manner to ensure that any biases were consistent across the entire dataset. JSpecies1.2.1 was used to analyze these contig sets for the ANI and tetramer usage patterns, using default parameters (Richter and Rosselló-Móra, 2009).

### Comparative Genomics

All of the orthologous pairs between *Pantoea* test genomes were identified by Pan Genome Analysis Pipeline (PGAP) (Zhao et al., 2012). The common dataset of shared genes among test strains was defined as their core genome. The total set of genes with test genomes was defined as the pan genome. The set of genes in each strain not shared with other strains was defined as the unique genes. The details of the strains used are listed in Supplementary Table S1.

### Construction of the Recombinant Plasmids and *Escherichia coli* Strains

A 3.86 kb *Bam*HI-*Xba*I DNA fragment containing the complete *ars1* cluster of *Pantoea stewartii* S301 (promoter region, 342 bp

<sup>1</sup> <http://www.bacterio.net/pantoea.html>

**TABLE 1 | Genes involved in arsenic resistance and transformation.**

Function	Gene	Protein
<b>Arsenate reduction</b>	<i>arsR</i>	Transcriptional regulator ArsR
	<i>arsB</i>	As(III) efflux pump protein
	<i>arsC</i>	As(V) reductase ArsC
	<i>arsH</i>	Putative flavoprotein
	<i>arsA</i>	As(III) active ATPase
	<i>arsD</i>	As metallochaperone
	<i>arsO</i>	Monooxygenase
	<i>arsT</i>	Putative thioredoxin system
	<i>arsX</i>	Unknown function
	<i>arsN</i>	Acetyltransferases
<b>Respiratory reduction</b>	<i>arrA</i>	Large subunit of respiratory As(V) reductase
	<i>arrB</i>	Small subunit of respiratory As(V) reductase
<b>Arsenite oxidation</b>	<i>aioA</i>	Large subunit of As oxidase
	<i>aioB</i>	Small subunit of As oxidase
	<i>aioD</i>	Biosynthesis protein A with molybdenum cofactor
	<i>aioX</i>	Phosphonate-binding periplasmic protein
	<i>aioS</i>	Histidine kinase for signal transduction
	<i>aioC</i>	Cytochrome c
<b>Arsenic methylation</b>	<i>arsM</i>	As(III) S-adenosylmethyltransferase

<sup>2</sup> <http://tree.bio.ed.ac.uk/software/figtree/>



upstream of the start codon ATG of *arsR*, the contiguous four genes *arsR1B1C1H1* and 281 bp upstream of the start codon ATG of *arsH* was PCR amplified with primers Ars1-F and Ars1-R (Supplementary Table S2). A 3.43 kb *Bam*HI-*Xba*I DNA fragment containing the complete *ars2* gene cluster of *P. agglomerans* Tx10 (a 280 bp region downstream of the stop codon TAA of *arsC2* and the contiguous ten genes *arsR2B2C2H2* and 328 bp downstream of the stop codon TAA of *arsH2*) was PCR amplified with primers Ars2-F and Ars2-R (Supplementary Table S2).

An 860 bp *Bam*HI-*Xba*I DNA fragment containing the complete *arsC1*-like gene of *P. stewartii* DC283 (promoter region, 221 bp upstream of the start codon ATG of *arsC1*-like gene, *arsC1*-like and 209 bp downstream of the stop codon TTA of *arsC1*-like gene) was PCR amplified with primers ArsC1-like-F and ArsC1-like-R (Supplementary Table S2). A 942 bp *Bam*HI-*Xba*I DNA fragment containing the complete *arsC2*-like gene of *P. stewartii* DC283 (promoter region, 265 bp upstream of the start codon ATG of *arsC2*-like gene, *arsC2*-like and 236 bp downstream of the stop codon TTA of *arsC2*-like gene) was PCR amplified with primers ArsC2-like-F and ArsC2-like-R (Supplementary Table S2).

The above PCR products were ligated to the *Bam*HI-*Xba*I site of plasmid pUC18, yielding plasmids pUC18-*ars1*, pUC18-*ars2*, pUC18-*arsC1*-like, and pUC18-*arsC2*-like. Then the plasmids were transferred to *E. coli* AW3110, yielding the recombinant *E. coli* AW3110-*ars1*, *E. coli* AW3110-*ars2*, *E. coli* AW3110-*arsC1*-like and *E. coli* AW3110-*arsC2*-like strains, respectively.

## Strains, Plasmids, and Culture Conditions

The strains and plasmids used in this work are summarized in Supplementary Table S3. *E. coli* and *Pantoea* strains were grown in LB medium (per liter contains: 10 g tryptone, 5 g yeast, and 10 g NaCl) or LB plates (LB medium with w/v 1.5% agar) at 30°C. When appropriate, antibiotics were added at the following concentration: 100 µg/mL ampicillin. Resistance to As species was tested by plating serial dilutions of cultures of each strain onto agar plates containing filtered sodium arsenate (Na<sub>3</sub>AsO<sub>4</sub>).

## RESULTS

### Genomic Features

To date, 26 species have been reported in genus *Pantoea* and strains of nine species (*P. ananatis*, *P. agglomerans*, *P. stewartii*, *P. vagans*, *P. dispersa*, *P. septica*, *P. rodasii*, *P. rwandensis*, and *P. anthophila*) have been sequenced<sup>3</sup>. To study the genetic traits and phylogenetic history of As resistance in genus *Pantoea*, 23 strains were chosen, containing two to three standard strains sequenced in each species and five unidentified strains (Supplementary Table S1). A summary of features for these 23 sequenced genomes is listed in Supplementary Table S1. The G+C contents of the 23 genomes range from 53.4 to 59.1. These genomes vary in size by approximately 1.6 mega-bases in average (ranging from 4.02

to 5.68 Mb) with coding sequence (CDS) numbers ranging from 3580 to 8894, indicating a substantial strain-to-strain variation.

### Strain-Specific and Core Genes

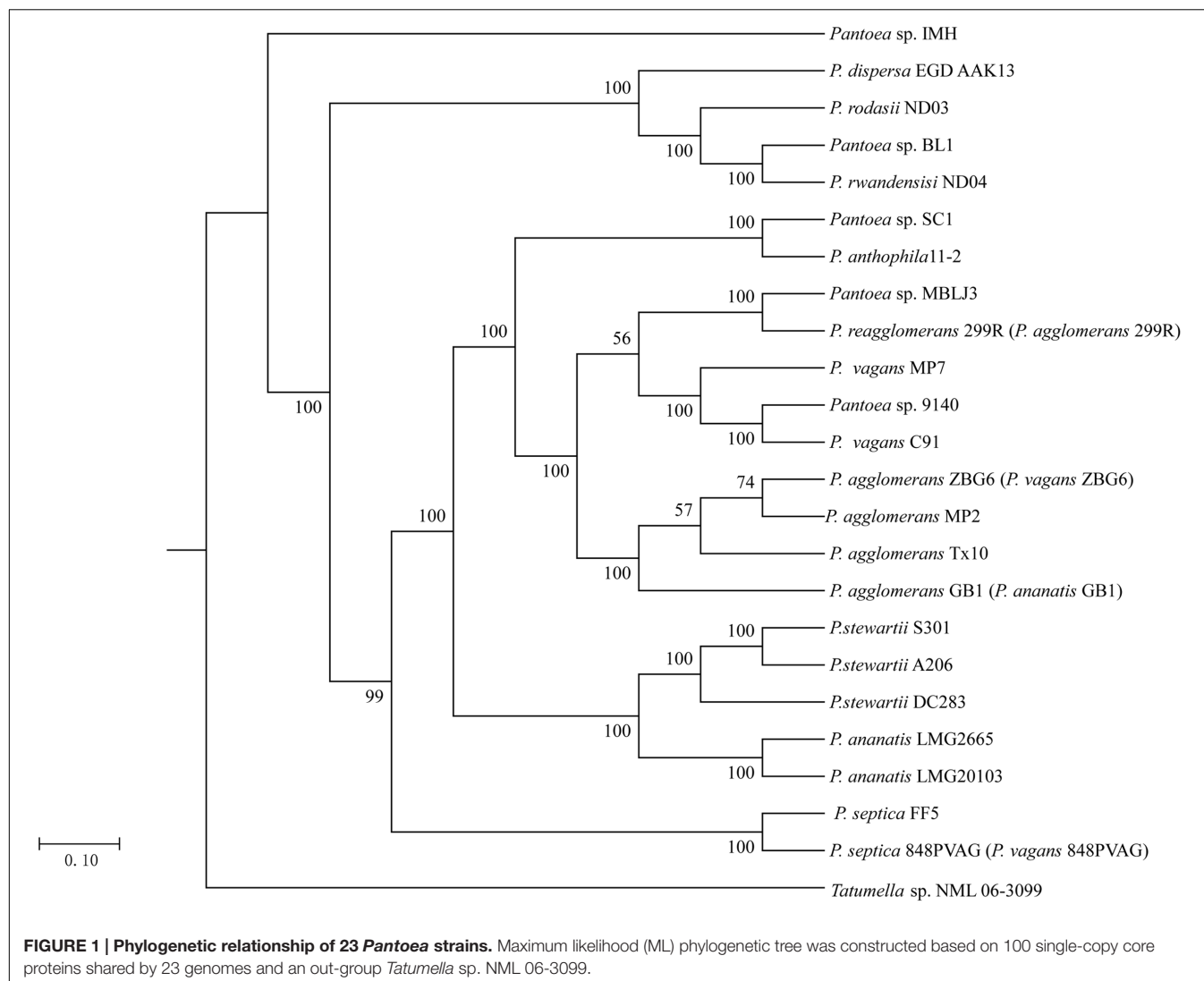
To reveal the genomic features specific to each strain, we identified all orthologous pairs between the tested *Pantoea* genomes using PGAP. Our analysis of the total of 23 genomes revealed that a pan genome contains 48,207 putative protein-coding genes in the genus *Pantoea*. Out of these 48,207 genes, 10,896 (22.6%) were represented in the specific genomes of *Pantoea* spp., suggesting some frequency of horizontal gene acquisition from other taxa. The number of specific genes ranges from 131 to 1,285, with the smallest encoded by *P. vagans* C9-1 and the largest identified in *P. agglomerans* Tx10 (Supplementary Figure S1). The cluster of orthologous groups (COG) assignments reveal that a higher proportion of strain-specific genes in most of the strains can be assigned to the K (transcription), L (DNA replication), and M (cell wall/membrane/envelope biogenesis) categories (Supplementary Figure S2).

In contrast to the pan-genome, the core genome of *Pantoea* spp. contains 1,994 putative protein-coding genes, which represents 38.8–56.1% of the repertoire of protein coding genes of each strain, illustrating a small degree of genomic diversity in this group of bacteria (Supplementary Figure S1). The genomic analysis agrees with the fact that *Pantoea* strains are consistent in morphological and physiological appearance. Furthermore, the COG assignment results show that these core genes are in different functional categories (Supplementary Figure S3). In fact, the percentage of genes in each functional category remains rather similar (with an average divergence of 8.6%). This is consistent with an earlier report that larger prokaryotic genomes preferentially accumulate genes directly or indirectly involved in metabolism (Konstantinidis and Tiedje, 2004). These genes support a broader metabolic diversity, which, in turn, would improve the ecological success of *Pantoea* under more diverse environmental conditions.

### Phylogenetic Analyses

To associate the distribution of As resistance genes in *Pantoea* spp. with their phylogenetic affiliation, we constructed the phylogeny tree of the 23 *Pantoea* spp. based on 16S rRNA gene sequences using NJ methods rooted by *Tatumella* sp. NML 06-3099 (Supplementary Figure S9). This phylogenetic tree showed that the strains in the same species reported were grouped together except strain 848PVAG. At the same time, we constructed the phylogeny of the 23 genomes based on concatenation of the 100 core genes that are present as single copies in a genome using the ML method and rooted by *Tatumella* sp. NML 06-3099 (Figure 1). The phylogenetic trees, inferred using BI and NJ methods (Supplementary Figures S4, S5), were congruent with the ML phylogenetic tree. These trees show that some strains in different species reported were grouped together, such as, FF5 and 848PVAG, ZBG6, GB1, MP2, and Tx10. The phylogeny

<sup>3</sup><http://www.ncbi.nlm.nih.gov/genome/?term=pantoea>



based on the 100 core genes that are present as single copies in a genome showed a good correlation with that of 16S rRNA gene sequences, except for three strains ZBG6, GB1, and 299R. These results suggested that there were mistakes in the classification of *Pantoea* spp.. Further identification of the phylogenetic status of these strains was carried out as follows.

The information gained from the phylogenetic analysis provides an important depiction of the evolutionary relationship between different strains, but it does not translate directly into the overall similarity of the genomes, which is usually determined through the DNA-DNA hybridization (DDH). Herein, ANI approach was used to overcome the difficulty of conventional laboratory-based DDH in evaluating the genomic similarity of bacteria (Richter and Rosselló-Móra, 2009). The ANI results justified the conclusion of phylogenetic analysis. As shown in **Figure 2**, 23 strains were classified into 12 species based on their ANI  $\geq 96\%$ . For examples, LMG2665 and LMG20103 resulted in a higher ANI (99.3%), suggesting that they belong to the same

species (*P. ananatis*). Strain 9140 and C91 resulted in a higher ANI (98.6%), suggesting that they belong to the same species (*P. vagans*). It was noteworthy that *Pantoea* sp. IMH represented a novel species for the ANI  $\geq 96\%$  between IMH and other strains.

Strains MP2, Tx10, GB1, and ZBG6 which grouped together were identified as strains of *P. agglomerans*. Meanwhile, this result confirms the synonymy of *P. FF5* and 848 PVAG (*P. vagans*), and suggests that 299R is not a member of species *P. agglomerans*. In agreement with the phylogenetic analysis, our ANI results indicate that there are mistakes in the classification of strain 299R, 848PVAG, GB1, and ZBG6. This mis-classification was also reported in other genus and generally corrected with the advance in technology (Goris et al., 2007). To associate the distribution of As-related genes with their phylogenetic affiliation, in this article below we renamed strain 299R to *P. reaggglomerans* 299R (*P. agglomerans* 299R), 848PVAG to *P. septica* 848PVAG (*P. vagans* 848PVAG), GB1 to *P. agglomerans* GB1 (*P. ananatis* GB1), and ZBG6 to *P. agglomerans* ZBG6 (*P. vagans* ZBG6).

■ = ANI ≥ 96%

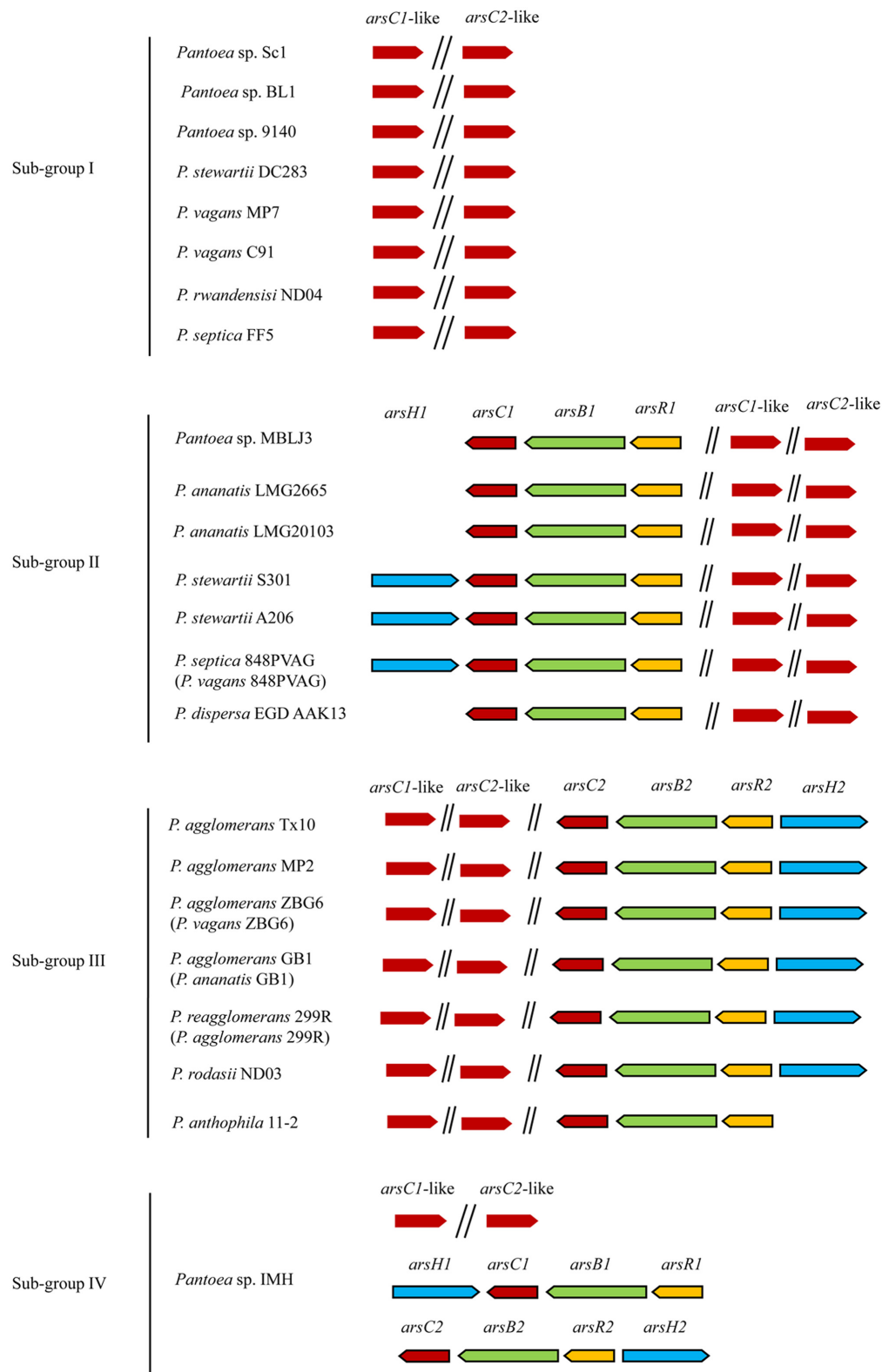
	<i>Tatumella saanichensis</i> NML 06-3099	<i>Pantoea stewartii</i> S301	<i>Pantoea stewartii</i> A206	<i>Pantoea stewartii</i> DC283	<i>P. agglomerans</i> ZBG6 ( <i>P. vagans</i> ZBG6)	<i>Pantoea agglomerans</i> MP2	<i>Pantoea agglomerans</i> Tx10	<i>P. agglomerans</i> GB1 ( <i>P. ananatis</i> GB1)	<i>P. septica</i> 848PVAG ( <i>P. vagans</i> 848PVAG)	<i>Pantoea septica</i> FF5	<i>Pantoea</i> sp. BL1	<i>Pantoea</i> sp. 9140	<i>Pantoea rwandensis</i> ND04	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140	<i>Pantoea</i> sp. 9140
<i>Pantoea ananatis</i> LMG 2665																								
<i>Pantoea ananatis</i> LMG 20103																								
<i>Pantoea anthophila</i> 11-2																								
<i>Pantoea</i> sp. Sc1																								
<i>P. reaggglomerans</i> 299R ( <i>P. agglomerans</i> 299R)																								
<i>Pantoea</i> sp. MBLJ3																								
<i>Pantoea dispersa</i> EGD-AAK13																								
<i>Pantoea</i> sp. IMH																								
<i>Pantoea rodarii</i> ND03																								
<i>Pantoea vagans</i> C9-1																								
<i>Pantoea vagans</i> MP7																								
<i>Pantoea</i> sp. 9140																								
<i>Pantoea rwandensis</i> ND04																								
<i>Pantoea</i> sp. BL1																								
<i>Pantoea septica</i> FF5																								
<i>P. septica</i> 848PVAG ( <i>P. vagans</i> 848PVAG)																								
<i>P. agglomerans</i> GB1 ( <i>P. ananatis</i> GB1)																								
<i>Pantoea agglomerans</i> Tx10																								
<i>Pantoea agglomerans</i> MP2																								
<i>P. agglomerans</i> ZBG6 ( <i>P. vagans</i> ZBG6)																								
<i>Pantoea stewartii</i> DC283																								
<i>Pantoea stewartii</i> A206																								
<i>Pantoea stewartii</i> S301																								
<i>Tatumella saanichensis</i> NML 06-3099																								

**FIGURE 2 | Average Nucleotide Identity (ANI) (%) based on whole genome alignments.** ANI values are colored red according to historical species cutoff value ( $\geq 96\%$ ). Strains in one species are marked out the same color.

## Distribution and Organization of As-Related Genes in *Pantoea* Genomes

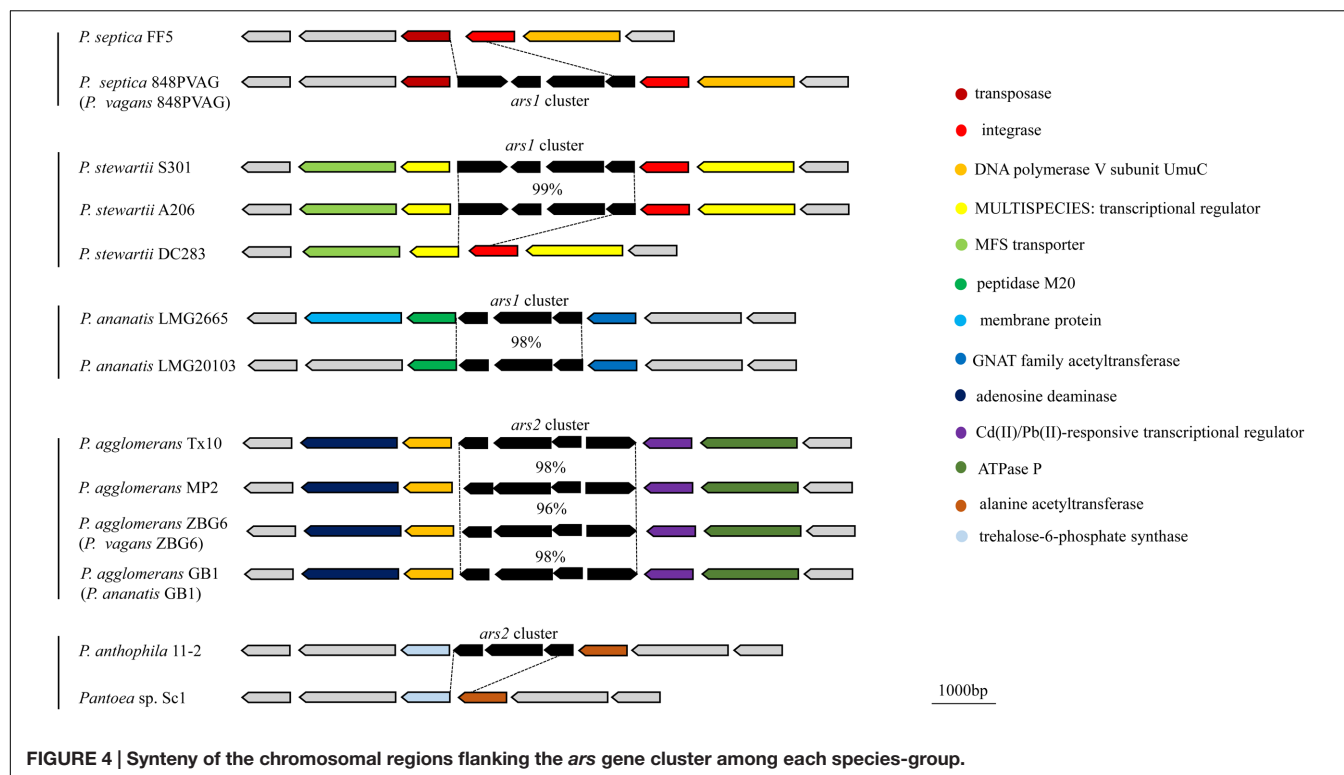
Only As resistance genes (*ars* genes) including *arsR*, *arsB*, *arsC*, and *arsH* were detected in most *Pantoea* genomes (Supplementary Table S4 and Figure 3). The *arsC* gene encoding arsenate reductase is involved in the transformation of As(V) to As(III), which is then excreted by the As efflux pump ArsB encoded by the *arsB* gene. Nevertheless, *aio*, *arr*, and *arsM* were not found in *Pantoea* genomes, suggesting that cytoplasmic As(V) reduction and As(III) extrusion are the As resistance strategy used in genus *Pantoea* spp.. This mechanism benefits the bacteria itself, though it enhances the toxicity to the surrounding environment.

The *ars* genes in a genome are prone to group together as *ars* clusters (*arsRBC* and *arsRBCH*). Although comparison of the COG assignments of 23 genomes revealed that the DNA sequences between homologous genes within these *ars* clusters are conserved, some variations exist in DNA sequences, which can be divided into two sub-groups (*ars1* and *ars2*) (Figure 3). Unlike the two *ars* clusters in *Pantoea* sp. IMH, only one *ars* cluster, either *ars1* or *ars2*, was observed in other strains (Supplementary Table S2 and Figure 3). The *ars* gene clusters generally exhibited more than 80% identity within each sub-group and about 54% identity between two sub-groups. Actually, the *ars* clusters were not detected in eight strains including Sc1, BL1, 9140, DC283, MP7, C91, ND04, and FF5. Moreover, two



**FIGURE 3 | Distribution and organization of *ars* genes and *arsC*-like genes in 23 *Pantoea* strains.** *arsC*, *arsB*, *arsR*, *arsH*, and *arsC*-like genes are marked with different colors. There are only *arsC*-like genes in Sub-group I, *arsR1B1C1* or *arsR1B1C1H1* in Sub-group II, *arsR2B2C2* or *arsR2B2C2H2* in Sub-group III, and both *arsR1B1C1H1* and *arsR2B2C2H2* in Sub-group IV.





*arsC*-like genes with only 25% homology (*arsC1*-like and *arsC2*-like) were found in the 23 genomes. Based on the different *ars* genes distributions, the 23 strains were categorized into four sub-groups and discussed as follows. The overall distribution and organization of As resistance genes in 23 *Pantoea* strains are summarized in **Figure 3**.

## Evolution and the Origin of *ars* Clusters

The distribution and organization of *ars* genes in *Pantoea* raise a question as to their evolution. The deviant G+C content is used as a detect method of HGT (Ochman et al., 2000; Xie et al., 2014). We detected the G+C content of *ars* clusters and their corresponding genomes. The results showed that the G+C contents of the *ars1* clusters are higher than those of the genomes in *Pantoea* strains (56.3–57.8 vs. 53.4–54.7) except *P. septica* 848PVAG (*P. vagans* 848PVAG) and *P. dispersa* EGD-AAK13; the G+C contents of the *ars2* clusters are lower than those of the genomes in *Pantoea* strains (50.6–52.4 vs. 53.7–58.8), showing variation of G+C content between clusters and the corresponding genomes. These results indicated that these *ars* clusters may be acquired in *Pantoea* strains by HGT (Supplementary Table S4 and Figure S6). To further elucidate the evolution of the *ars* gene clusters, we compared the chromosomal regions flanking the *ars* gene clusters among the 23 *Pantoea* strains and found that the genes in the upstream and downstream regions were conserved among strains of the same species (**Figure 4**). For example, the DNA polymerase V subunit UmuC gene and adenosine deaminase gene in the upstream and Cd(II)/Pb(II)-responsive transcriptional regulator gene and ATPase P gene in the

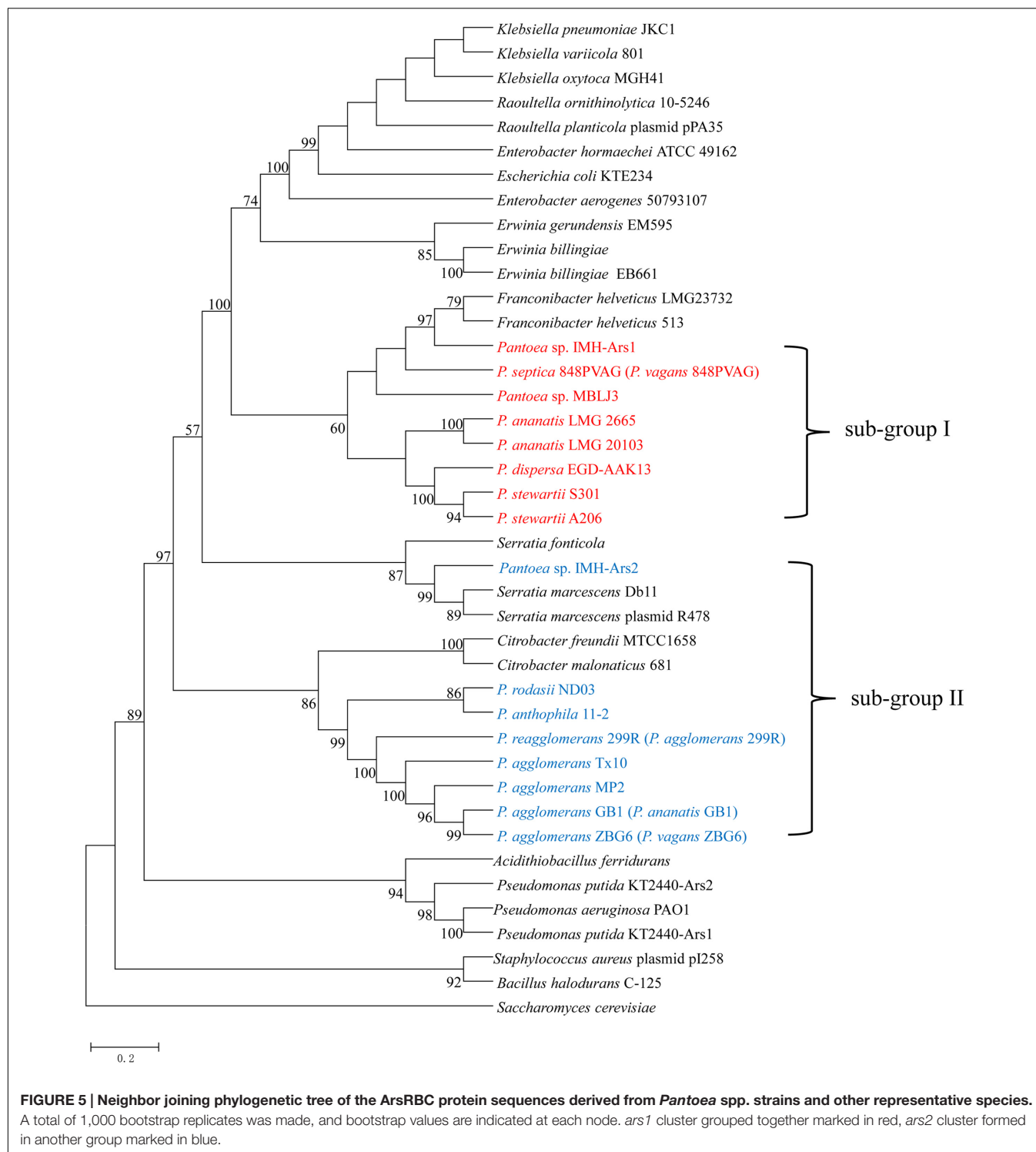
downstream are conserved for the *ars* clusters in the strain Tx10, MP2, ZBG6, and GB1 within the species *P. agglomerans* strains. The same species strains share the same insertion sites, whereas the different species' strains result in different insertion sites, suggesting that *ars* clusters may be acquired more than once.

Interestingly, as shown in **Figure 4**, the flanking regions of the *ars* gene clusters in strain *P. stewartii* S301 and *P. stewartii* A206 were homologous to the corresponding regions of strain *P. stewartii* DC283; the same phenomenon was found in strain *P. septica* 848PVAG (*P. vagans* 848PVAG) and *P. septica* FF5, and strain *P. anthophila* 11-2 and *Pantoea* sp. Sc1. This result suggests that *ars* clusters may be lost in *P. stewartii* DC283, *P. septica* FF5, and *Pantoea* sp. Sc1.

To gain insights into the origin of *ars* genes clusters in *Pantoea*, a NJ phylogenetic tree was constructed based on the ArsRBC protein sequences. As shown in **Figure 5**, the strains possessing *ars1* and *ars2* clusters form separate groups. Notably, the phylogeny reveals that the *ars1* and *ars* clusters of *F. helveticus* were sister groups, and *ars2* grouped to *ars* clusters of *S. marcescens* and *C. freundii*. These results imply that the *ars1* cluster may be acquired via HGT from *F. helveticus*, and *ars2* from *S. marcescens* and *C. freundii* in early evolutionary history.

## Two *arsC*-Like Genes in *Pantoea*

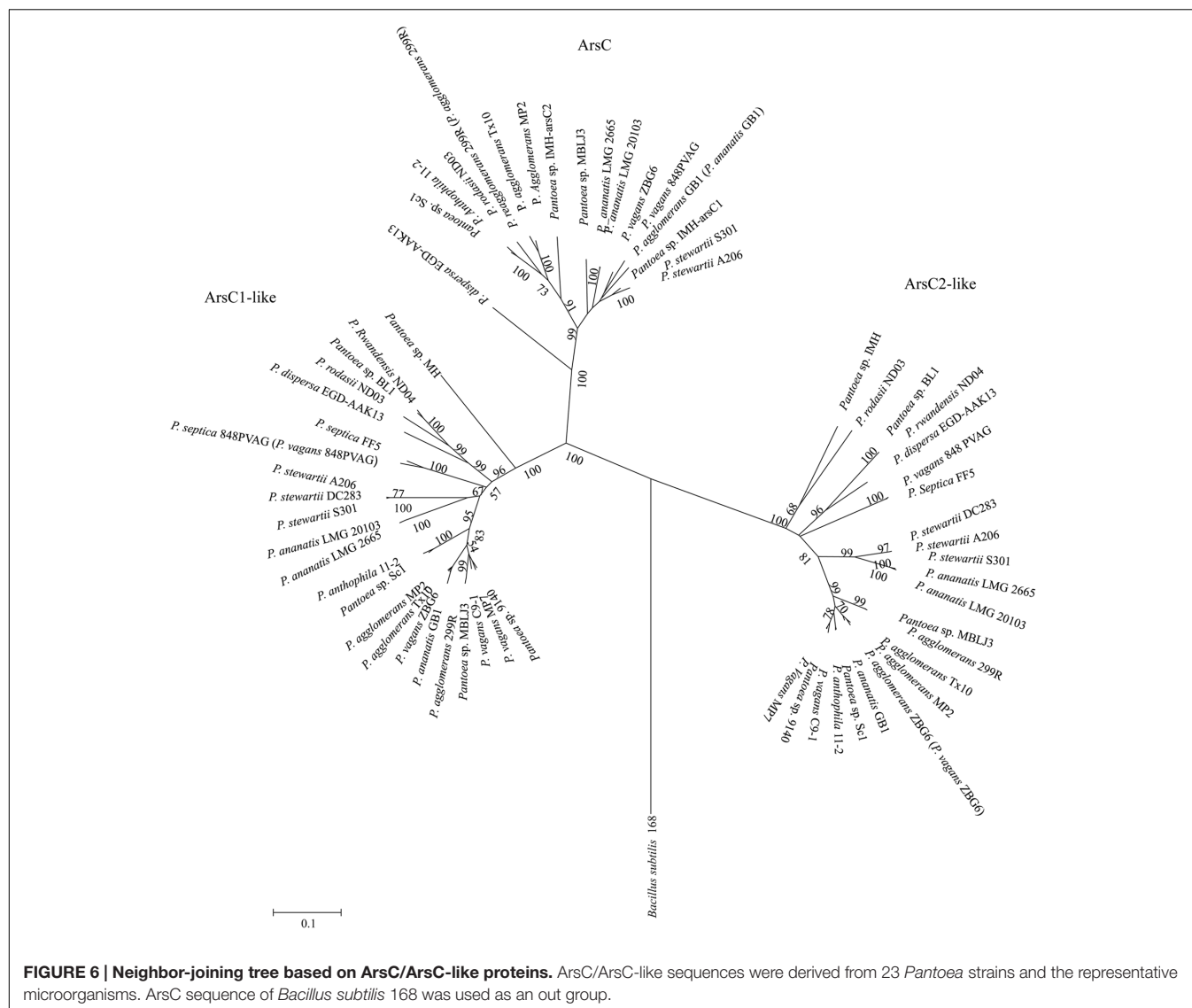
Our studies reveal that two *arsC*-like genes (*arsC1*-like and *arsC2*-like) are found in the 23 genomes with just 25% homology (**Figure 3**). Our phylogenetic analysis showed that the ArsC-like sequences formed distinct groups, which were clearly divergent from conventional arsenate reductase (**Figure 6**). It was reported



that Cys-12, Arg-60, Arg-94, and Arg-107 were four conserved residues of the ArsC protein in the process of arsenic resistance (Gladysheva et al., 1996). Cys-12 was identified as a catalytic residue and was activated by nearby residues Arg-60, Arg-94, and Arg-107 (Martin et al., 2001). Alignment analysis of *arsC* and *arsC*-like genes shows that Cys-12 and Arg-94 residues were

conserved, but residues Arg-107 and Arg-60 in two ArsC-like proteins were not conserved (Supplementary Figure S8). These results suggest that these two *arsC*-like genes are not involved in the As resistance.

To explore the evolution of these two *arsC*-like genes in *Pantoea*, molecular phylogenetic analysis, molecular

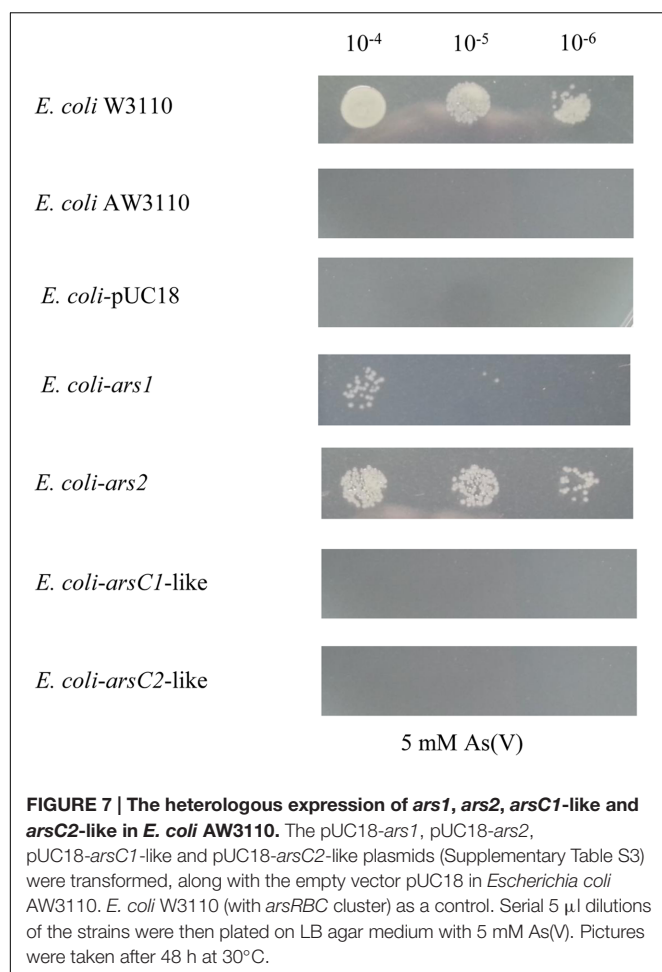


conservation, and linear representation analysis were used (Rice and Lampson, 1995; Nelson et al., 1999; Brochier-Armanet and Forterre, 2006; Dagan et al., 2008). The comparative analysis showed that the two *arsC*-like genes are conserved in all of the 23 (Supplementary Figure S7). Phylogenetic analysis showed that *arsC1*-like and *arsC2*-like genes were grouped together, respectively (Figure 6). These results suggested that the two *arsC*-like genes evolved with a possible evolutionary scenario of that there is a common ancestor. Further, we compared the flanking regions of the two *arsC*-like genes. Interestingly, two genes in the upstream (the uracil phosphoribosyl transferase genes and uracil/xanthine transporter genes) and two genes in the downstream (sulfur reduction protein DsrE and GntR family transcriptional regulator genes) are conserved for *arsC1*-like genes. The two genes in the upstream (DNA-binding response regulator genes and multidrug efflux RND transporter permease genes) and two genes in the downstream (succinyl-diaminopimelate desuccinylase genes and membrane protein

genes) are also conserved for *arsC2*-like genes (Supplementary Figure S7). This observation also suggests that *arsC1*-like and *arsC2*-like genes were the vertical genetic genes in the genus *Pantoea*. Possibility, they may have been the main As resistance contributors in early times and later had evolved with deviance.

## Functional Analysis of *ars* Gene and *arsC*-Like Genes

To verify that the *ars* gene clusters are the contributors to the As resistance, the *ars1* cluster with its promoter from *P. stewartii* S301, a representative strain with the *ars1* cluster, and the *ars2* cluster with its promoter from *P. agglomerans* Tx10, a representative strain with the *ars2* cluster, were PCR amplified and then ligated into vector pUC18 and further transferred to *E. coli* AW3110 (without any As resistance genes). The growth of the yielded recombinant *E. coli* strains *E. coli*-*ars1* and *E. coli*-*ars2*, was tested in 5 mM concentration As(V). As shown in Figure 7, both *E. coli*-*ars1* and *E. coli*-*ars2* survived in 5 mM As(V), and



*E. coli-ars1* grew better than *E. coli-ars2*. This result suggests that both the *ars1* and *ars2* clusters enabled *E. coli* AW3110 to resist As, and *ars1* seemed to have a more effective As resistance capability than *ars2*.

To test the functions of *arsC1*-like and *arsC2*-like genes, the *arsC1*-like gene and *arsC2*-like gene with their promoters from *P. stewartii* DC283 were PCR amplified and then ligated into vector pUC18 and further transferred to *E. coli* AW3110. As shown in **Figure 7**, neither the *arsC1*-like nor *arsC2*-like gene enables *E. coli* AW3110 to resist As. In line with the alignment result, the function analysis demonstrates that these two *arsC*-like genes do not contribute to As resistance.

## DISCUSSION

*Pantoea* is a genus with 26 members identified by DDH, a gold standard for prokaryotic species identification. However, laboratory-based DDH results may be irreproducible, and vary depending on the reannealing temperature (Gevers et al., 2005). With the rapid development in technology and decline in sequencing cost, promising new measurements such as ANI are being developed to evaluate the genomic similarity of bacteria (Richter and Rosselló-Móra, 2009). In this study, we identified

the 23 *Pantoea* spp. phylogenetic status using ANI, together with phylogenetic trees based on concatenated sequences of the 100 core genes (**Figures 1, 2**). Our results showed that strain 299R, 848PVAG, GB1, and ZBG6 were misnamed. We reclassified strain 299R to *P. reaggglomerans* 299R (*P. agglomerans* 299R), 848PVAG to *P. septica* 848PVAG (*P. vagans* 848PVAG), GB1 to *P. agglomerans* GB1 (*P. ananatis* GB1), and ZBG6 to *P. agglomerans* ZBG6 (*P. vagans* ZBG6). Our study provided data from genus *Pantoea* with a complex and controversial taxonomy and demonstrated the accuracy of a bioinformatics approach, such as ANI, to identify new species and to correct erroneous identifications from previous studies.

A previous study suggested that the *ars* system is a widespread As resistance mechanism (Páezespino et al., 2009). *Pantoea* sp. IMH was found to resist As by means of *ars* clusters. *arsRBC* is located on the large universal *Pantoea* plasmids of four strains including *P. agglomerans* E325, *P. agglomerans* MP2, *P. eucalyptia*B, and *P. anthophila* Sc1 (Maayer et al., 2012). This information leads to the hypothesis that plasmids may be involved in the evolution of As resistance mechanism by *ars* genes in *Pantoea* spp. However, there are untouched questions such as what are the mechanisms of the other *Pantoea* spp. and what is the evolutionary history of genetic elements involved in the As resistance? To answer these questions, we collected the genome sequences of 23 strains in nine species in NCBI (*P. ananatis*, *P. agglomerans*, *P. stewartii*, *P. vagans*, *P. dispersa*, *P. septica*, *P. rodasii*, *P. rwandensis*, and *P. anthophila*). The sequencing results provided us with mass genomic information to detect the presence and the locations of As-related genes in *Pantoea* spp. Our study for the first time systematically analyzed the As resistance genes and revealed the As resistance traits in genus *Pantoea*. Our research provided the definitive evidence that that As resistance strategy in *Pantoea* spp. only involved the detoxification mechanism through *ars* clusters, not the respiratory reduction mechanism through *arr* clusters. This detoxification strategy was obtained by HGT. This conclusion can likely to be extended to most bacteria. We speculate that evolutionarily ancient microbes were exposed to As surroundings on ancient earth (Oremland et al., 2009). To overcome the As-induced selection pressure, microbes evolved *ars* genes in their genomes for survival by HGT. Therefore, *ars* has very early origins and represents a widespread As resistance mechanism.

Two scattered *arsC*-like genes exist in each genome of the 23 *Pantoea* strains, but they exhibited no functional As resistance. It is rare for *arsC*-like genes to show no As resistance capabilities (Butcher et al., 2000; Saltikov and Newman, 2003). Compared to functional protein ArsC, residues Arg-107 and Arg-60 of ArsC-like protein were variant (Supplementary Figure S8). We speculate that in early times, the ancestor of *Pantoea* spp. evolved the *arsC* gene to resist As, but later evolved with deviance during adaption to As-free niches, and thus retained non-functional *arsC*-like genes in some genomes.

The *ars* genes are abundant and tend to organize in typical *arsRBC* cluster structures (**Figure 3**). Apart from these operons, *arsRBCH* operons are widely observed. In genus *Pantoea*, these kinds of structures were anticipated, for these strains descended from a recent common ancestor. Our study suggests



that *ars* clusters may be acquired by HGT from *F. helveticus*, *S. marcescens*, and *C. freundii* strains. This is consistent with recent literature showing that bacterial As resistance and transformation was a trait acquired via HGT, driven by adaptation to habitats containing As (Cai et al., 2009; Villegas-Torres et al., 2011). Interestingly, *ars* clusters are absent in some strains, suggesting that some microbes may have lost their As resistance genes during adaption to As-free niches. In addition, the number of As resistance genes in strains isolated from As-rich environments is much higher than in strains from other environments (Macur et al., 2004; Sutton et al., 2009). Compared to the evolutionary pattern of *ars* operons (Rosen, 1999), the evolution of As resistance genes (*ars* clusters) in *Pantoea* spp. involves a mix of HGT and loss, providing insight into the complex evolutionary history of As resistance.

## AUTHOR CONTRIBUTIONS

LW and CJ conceived and designed the study. LW performed the laboratory work and data analysis. LW, JW, and CJ drafted the tables and figures, and prepared the main manuscript.

## REFERENCES

- Arsène-Ploetze, F., Koechler, S., Marchal, M., Coppée, J. Y., Chandler, M., Bonnefoy, V., et al. (2010). Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet.* 6:e1000859. doi: 10.1371/journal.pgen.1000859
- Brochier-Armanet, C., and Forterre, P. (2006). Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* 2, 83–93.
- Butcher, B. G., Deane, S. M., and Rawlings, D. E. (2000). The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. *Appl. Environ. Microbiol.* 66, 1826–1833. doi: 10.1128/AEM.66.5.1826-1833.2000
- Cai, L., Liu, G., Rensing, C., and Wang, G. (2009). Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. *BMC Microbiol.* 9:4. doi: 10.1186/1471-2180-9-4
- Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., and Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using aeromonas as a test case. *MBio* 5:e2136-14. doi: 10.1128/mBio.02136-14
- Dagan, T., Artzyrandrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10039–10044. doi: 10.1073/pnas.0800679105
- Gavini, F., Mergaert, J., Bej, A., Mielcarek, C., Izard, D., Kersters, K., et al. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int. J. Syst. Evol. Microbiol.* 39, 337–345.
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739. doi: 10.1038/nrmicro1236
- Ghosh, D., Bhadury, P., and Routh, J. (2014). Diversity of arsenite oxidizing bacterial communities in arsenic-rich deltaic aquifers in West Bengal, India. *Front. Microbiol.* 5:602. doi: 10.3389/fmicb.2014.00602
- Gladysheva, T., Liu, J., and Rosen, B. P. (1996). His-8 lowers the pKa of the essential Cys-12 residue of the ArsC arsenate reductase of plasmid R773. *J. Biol. Chem.* 271, 33256–33260. doi: 10.1074/jbc.271.52.33256
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0
- Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3160–3165. doi: 10.1073/pnas.0308653100
- Levin, D., and Tal, A. (2003). Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium. *J. Bacteriol.* 185, 135–141. doi: 10.1186/1471-2164-11-709
- Maayer, P. D., Chan, W. Y., Blom, J., Venter, S. N., Duffy, B., Smits, T. H. M., et al. (2012). The large universal *Pantoea* plasmid LPP-1 plays a major role in biological and ecological diversification. *BMC Genomics* 13:625. doi: 10.1186/1471-2164-13-625
- Macur, R. E., Jackson, C. R., Botero, L. M., McDermott, T. R., and Inskeep, W. P. (2004). Bacterial populations associated with the oxidation and reduction of arsenic in an unsaturated soil. *Environ. Sci. Technol.* 38, 104–111.
- Martin, P., DeMel, S., Shi, J., Gladysheva, T., Gatti, D. L., Rosen, B. P., et al. (2001). Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. *Structure* 9, 1071–1081. doi: 10.1016/S0969-2126(01)00672-4
- Meng, M., Sun, W. Q., Geelhaar, L. A., Kumar, G., Patel, A. R., Payne, G. F., et al. (1995). Denitration of glycerol trinitrate by resting cells and cell extracts of *Bacillus thuringiensis/cereus* and *Enterobacter agglomerans*. *Appl. Environ. Microbiol.* 61, 2548–2553.
- Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., et al. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329. doi: 10.1038/20601
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500
- Oremland, R. S., Saltikov, C. W., Wolfe-Simon, F., and Stolz, J. F. (2009). Arsenic in the evolution of earth and extraterrestrial ecosystems. *Geomicrobiol. J.* 26, 522–536. doi: 10.1080/01490450903102525
- Páezespino, D., Tamames, J., De, L. V., and Cánovas, D. (2009). Microbial responses to environmental arsenic. *Biometals* 22, 117–130. doi: 10.1007/s10534-008-9195-y
- Qin, J., Rosen, B. P., Zhang, Y., Wang, G., Franke, S., and Rensing, C. (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2075–2080. doi: 10.1073/pnas.0506836103
- Rezzonico, F., Smits, T. H., Montesinos, E., Frey, J. E., and Duffy, B. (2009). Genotypic comparison of *Pantoea agglomerans* plant and clinical strains. *BMC Microbiol.* 9:204. doi: 10.1186/1471-2180-9-204

## ACKNOWLEDGMENTS

We acknowledge the financial support of the National Basic Research Program of China (2015CB932003), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB14020302), and the National Natural Science Foundation of China (41373123, 41425016, 41503094, and 21321004). We thank Yongguan Zhu for the strain *E. coli* AW3110.

## DATA ACCESSIBILITY

The NCBI accession numbers of 23 draft genome sequences of *Pantoea* are listed in Supplementary Table S1.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00471/full#supplementary-material>

- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOS: the european molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)00204-2
- Rice, S. A., and Lampson, B. C. (1995). Phylogenetic comparison of retron elements among the myxobacteria: evidence for vertical inheritance. *J. Bacteriol.* 177, 37–45. doi: 10.1128/jb.177.1.37-45.1995
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Ronquist, F., Teslenko, M., van, der Mark P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Rosen, B. P. (1999). Families of arsenic transporters. *Trends Microbiol.* 7, 207–212.
- Saltikov, C. W., and Newman, D. K. (2003). Genetic identification of a respiratory arsenate reductase. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10983–10988. doi: 10.1073/pnas.1834303100
- Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17, 1246–1247. doi: 10.1093/bioinformatics/17.12.1246
- Smith, A. H., Lopipero, P. A., Bates, M. N., and Steinmaus, C. M. (2002). Public health. Arsenic epidemiology and drinking water standards. *Science* 296, 2145–2146. doi: 10.1126/science.1072896
- Sutton, N. B., van, der Kraan G. M., van, Loosdrecht M. C., Muyzer, G., Bruining, J., and Schotting, R. J. (2009). Characterization of geochemical constituents and bacterial populations associated with As mobilization in deep and shallow tube wells in Bangladesh. *Water Res.* 43, 1720–1730. doi: 10.1016/j.watres.2009.01.006
- Tamura, K. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* Chater 2:Unit 2.3. doi: 10.1002/0471250953.bi0203s00
- Tian, H., and Jing, C. (2014). Genome sequence of the aerobic arsenate-reducing bacterium *Pantoea* sp. strain IMH. *Genome Announc.* 2:e267–14. doi: 10.1128/genomeA.00267-14
- Villegas-Torres, M. F., Bedoya-Reina, O. C., Salazar, C., Vives-Florez, M. J., and Dussan, J. (2011). Horizontal *arsC* gene transfer among microorganisms isolated from arsenic polluted soil. *Int. Biodeter. Biodegradation* 65, 147–152. doi: 10.1016/j.ibiod.2010.10.007
- Wang, L., Zhuang, X., Zhuang, G., and Jing, C. (2016). Arsenic resistance strategy in *Pantoea* sp. IMH: organization, function and evolution of *ars* genes. *Sci. Rep.* 6:39195. doi: 10.1038/srep39195
- Wu, Q., Du, J., Zhuang, G., and Jing, C. (2013). *Bacillus* sp. SXB and *Pantoea* sp. IMH, aerobic As(V)-reducing bacteria isolated from arsenic-contaminated soil. *J. Appl. Microbiol.* 114, 713–721. doi: 10.1111/jam.12093
- Xie, J. B., Du, Z., Bai, L., Tian, C., Zhang, Y., Xie, J. Y., et al. (2014). Comparative genomic analysis of N<sub>2</sub>-fixing and non-N<sub>2</sub>-fixing *Paenibacillus* spp.: organization, evolution and expression of the nitrogen fixation genes. *PLoS Genet.* 10:e1004231. doi: 10.1371/journal.pgen.1004231
- Xu, C., Zhou, T., Kuroda, M., and Rosen, B. P. (1998). Metalloid resistance mechanisms in prokaryotes. *J. Biochem.* 123, 16–23.
- Zhang, L., and Birch, R. G. (1997). The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9984–9989.
- Zhao, C., Zhang, Y., Chan, Z., Chen, S., and Yang, S. (2015). Insights into arsenic multi-operons expression and resistance mechanisms in *Rhodopseudomonas palustris* CGA009. *Front. Microbiol.* 6:986. doi: 10.3389/fmicb.2015.00986
- Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. (2012). PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28, 416–418. doi: 10.1093/bioinformatics/btr655

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Wang, Wang and Jing. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Plasmid Replicons from *Pseudomonas* Are Natural Chimeras of Functional, Exchangeable Modules

Leire Bardaji<sup>1†</sup>, Maite Añorga<sup>1†</sup>, José A. Ruiz-Masó<sup>2</sup>, Gloria del Solar<sup>2</sup> and Jesús Murillo<sup>1\*</sup>

<sup>1</sup> Departamento de Producción Agraria, Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Pública de Navarra, Pamplona, Spain, <sup>2</sup> Molecular Biology of Gram-Positive Bacteria, Molecular Microbiology and Infection Biology, Centro de Investigaciones Biológicas (Consejo Superior de Investigaciones Científicas), Madrid, Spain

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

### Reviewed by:

Grzegorz Węgrzyn,  
University of Gdańsk, Poland  
Alan Leonard,  
Florida Institute of Technology, USA  
Jan Nesvera,  
Institute of Microbiology of the Czech  
Academy of Sciences, Czechia

### \*Correspondence:

Jesús Murillo  
jesus.murillo@unavarra.es

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 08 December 2016

**Accepted:** 25 January 2017

**Published:** 13 February 2017

### Citation:

Bardaji L, Añorga M, Ruiz-Masó JA,  
del Solar G and Murillo J (2017)  
Plasmid Replicons from  
*Pseudomonas* Are Natural Chimeras  
of Functional, Exchangeable Modules.  
*Front. Microbiol.* 8:190.  
doi: 10.3389/fmicb.2017.00190

Plasmids are a main factor for the evolution of bacteria through horizontal gene exchange, including the dissemination of pathogenicity genes, resistance to antibiotics and degradation of pollutants. Their capacity to duplicate is dependent on their replication determinants (replicon), which also define their bacterial host range and the inability to coexist with related replicons. We characterize a second replicon from the virulence plasmid pPsv48C, from *Pseudomonas syringae* pv. savastanoi, which appears to be a natural chimera between the gene encoding a newly described replication protein and a putative replication control region present in the widespread family of PFP virulence plasmids. We present extensive evidence of this type of chimerism in structurally similar replicons from species of *Pseudomonas*, including environmental bacteria as well as plant, animal and human pathogens. We establish that these replicons consist of two functional modules corresponding to putative control (REx-C module) and replication (REx-R module) regions. These modules are functionally separable, do not show specificity for each other, and are dynamically exchanged among replicons of four distinct plasmid families. Only the REx-C module displays strong incompatibility, which is overcome by a few nucleotide changes clustered in a stem-and-loop structure of a putative antisense RNA. Additionally, a REx-C module from pPsv48C conferred replication ability to a non-replicative chromosomal DNA region containing features associated to replicons. Thus, the organization of plasmid replicons as independent and exchangeable functional modules is likely facilitating rapid replicon evolution, fostering their diversification and survival, besides allowing the potential co-option of appropriate genes into novel replicons and the artificial construction of new replicon specificities.

**Keywords:** control and replication modules, chimeric replicons, gene co-option, Rep proteins, origin of replication, plasmid incompatibility, swapping of functional modules, virulence plasmids

## INTRODUCTION

Plasmids are extrachromosomal elements that colonize a vast majority of bacteria and other organisms, often carrying genes that confer an adaptive advantage to the host (del Solar et al., 1998; Jackson et al., 2011; Ruiz-Masó et al., 2015). Each cell can have from none to several plasmids of diverse sizes and copy numbers. Plasmids can readily acquire large amounts of foreign DNA from different sources and are transferred between distantly related organisms, including

prokaryotes and eukaryotes, which makes them a major contributor to the accessory gene pool and the most important agents in horizontal gene transfer (Halary et al., 2010; Jackson et al., 2011). Indeed, plasmids are responsible for the worldwide distribution of genes for resistance to antibiotics and other antimicrobials, rendering current strategies ineffective for the control of human, animal and plant diseases (Sundin, 2007; Jackson et al., 2011; Aviv et al., 2016; Johnson et al., 2016).

The basic replicon is the fundamental element for plasmid survival, ensuring timely duplication in coordination with cell division (Nordström, 1993; Summers, 1996; del Solar et al., 1998). Broadly, basic replicons consist of (i) a short *cis*-acting DNA sequence, the origin of replication, (ii) genes and structures involved in the control of replication and, for most plasmids, (iii) a gene coding a replication initiator (Rep) protein that recognizes the origin and promotes initiation of DNA replication. Plasmid replication is controlled by either directly repeated sequences (iterons) or by antisense RNAs, which can act alone or in coordination with a protein repressing transcription of the *rep* gene, and is tightly regulated so as to maintain the number of plasmid molecules in the cell within acceptable limits (Summers, 1996; del Solar and Espinosa, 2000). An immediate consequence of this is that plasmids sharing elements for replication or replication control cannot coexist in the same cell and are hence incompatible (Novick, 1987).

Replicons are highly diverse and can be grouped based on their general mechanism of replication, the function of their Rep proteins, their structure and genetic organization or their homology (del Solar et al., 1998; Lilly and Camps, 2015). Circular plasmids replicate by one of three general modes: rolling-circle, strand-displacement and theta-type mechanisms. According to their mode of replication initiation, the theta-type replicons have been grouped into four classes (A, B, C, and D) (Bruand et al., 1993). Class A theta replicons (e.g., R1, RK2, R6K, pSC101, pPS10, F and P) encode a Rep protein that binds to the origin and mediates melting of the duplex DNA. Class B (ColE1-like) replicons lack a *rep* gene, and melting of duplex DNA as well as synthesis of a pre-primer RNA for replication are achieved by bacterial RNA polymerase-mediated transcription. Class C (ColE2- and ColE3-like) replicons contain the smallest origins reported so far and encode a Rep primase protein that also mediates unwinding of the DNA (Itou et al., 2015). Finally, functioning of class D replicons (plasmids pAM $\beta$ 1, pIP501, and pSM19035 from Gram-positive bacteria) requires transcription across the origin and participation of a Rep protein in melting of the DNA and primer processing.

The gamma proteobacterial genus *Pseudomonas* comprises very diverse species, present in all kinds of environments, including significant human, animal and plant pathogens as well as species of outstanding biotechnological interest (Ramos, 2004). *Pseudomonas syringae* is one of the most relevant plant pathogenic bacteria in the world (Mansfield et al., 2012), and many strains carry one or more highly stable plasmids, ranging from a couple of kilobases to close to 1 Mb (Murillo and Keen, 1994; Sundin, 2007; Romanchuk et al., 2014). Most plasmids from *P. syringae*, and also various from many other *Pseudomonas* species, belong to the PFP (pPT23A-family plasmid) group

(Murillo and Keen, 1994; Sesma et al., 1998, 2000; Gibbon et al., 1999; Sundin, 2007). PFPs appear to originate from a common ancestor because they share homologous RepA-PFP replicons, which are related to the ColE2 class C theta replicons (Murillo and Keen, 1994; Sesma et al., 1998; Gibbon et al., 1999; Sesma et al., 2000; Sundin, 2007). ColE2 replicons contain a *rep* gene and an upstream region coding for a small antisense RNA, which is complementary to the 5'-nontranslated region of the *rep* mRNA and negatively controls its expression posttranscriptionally (Yasueda et al., 1994). Similarly, the RepA-PFP replicons consist of the *repA* replication initiator gene, which includes the putative vegetative origin of replication (Yagura et al., 2006), preceded by a short 5' sequence, containing diverse stem-and-loop (SaL) structures, which is probably involved in control of replication (Murillo and Keen, 1994; del Solar et al., 1998; Gibbon et al., 1999; Brantl, 2014). PFP plasmids have had tremendous evolutionary success, not only for their ubiquity across pseudomonads, but also because most *P. syringae* strains contain two to six coexisting PFP plasmids (Murillo and Keen, 1994; Sesma et al., 1998). This could be explained in part because they generally carry genes essential for the interaction with the plant host or for survival, fostering their frequent exchange among the bacterial population (Sesma et al., 2000; Vivian et al., 2001; Ma et al., 2007; Sundin, 2007; Sundin and Murillo, 2009). Additionally, their competitiveness among the bacterial plasmid pool might be enhanced by a replication machinery particularly adapted to their bacterial host. Notwithstanding, the coexistence of PFP plasmids in the same cell is difficult to explain because of their potential incompatibility (Novick, 1987; Sesma et al., 1998). In fact, PFP plasmids are generally incompatible with their cloned replicons (Murillo and Keen, 1994; Murillo et al., 1994; Sesma et al., 1998) although subcloning did not allow for the identification of the sequences responsible for this incompatibility within the replicon (Gibbon et al., 1999).

*Pseudomonas syringae* pv. savastanoi NCPPB 3335 contains three virulence PFP plasmids, pPsv48A (78 kb), pPsv48B (45 kb), and pPsv48C (42 kb), of which the smallest two appear to have originated by plasmid duplication and reorganization (Bardaji et al., 2011). Plasmid pPsv48C is essential for elicitation of disease symptoms in the plant host olive (*M. Añorga*, unpublished results), and is extremely stable (Bardaji et al., 2011). In this work, we identified a second replicon on pPsv48C, designated here as RepJ replicon, containing a putative replication control region homologous to that of the pPsv48C RepA-PFP replicon (Bardaji et al., 2011). We also show that these, and structurally similar replicons, consist of two functional modules corresponding to the putative control region (REx-C module) and the replication region (REx-R module). These modules are functionally separable, do not show specificity for each other, and are dynamically exchanged among replicons of four distinct families. Additionally, a REx-C module from pPsv48C conferred replication ability to a non-replicative *repJ* chromosomal homolog. Thus, the organization of plasmid replicons as independent and exchangeable functional modules is likely fostering their diversification and survival, besides allowing the potential co-option of appropriate genes into novel replicons and the artificial construction of new replicon specificities.



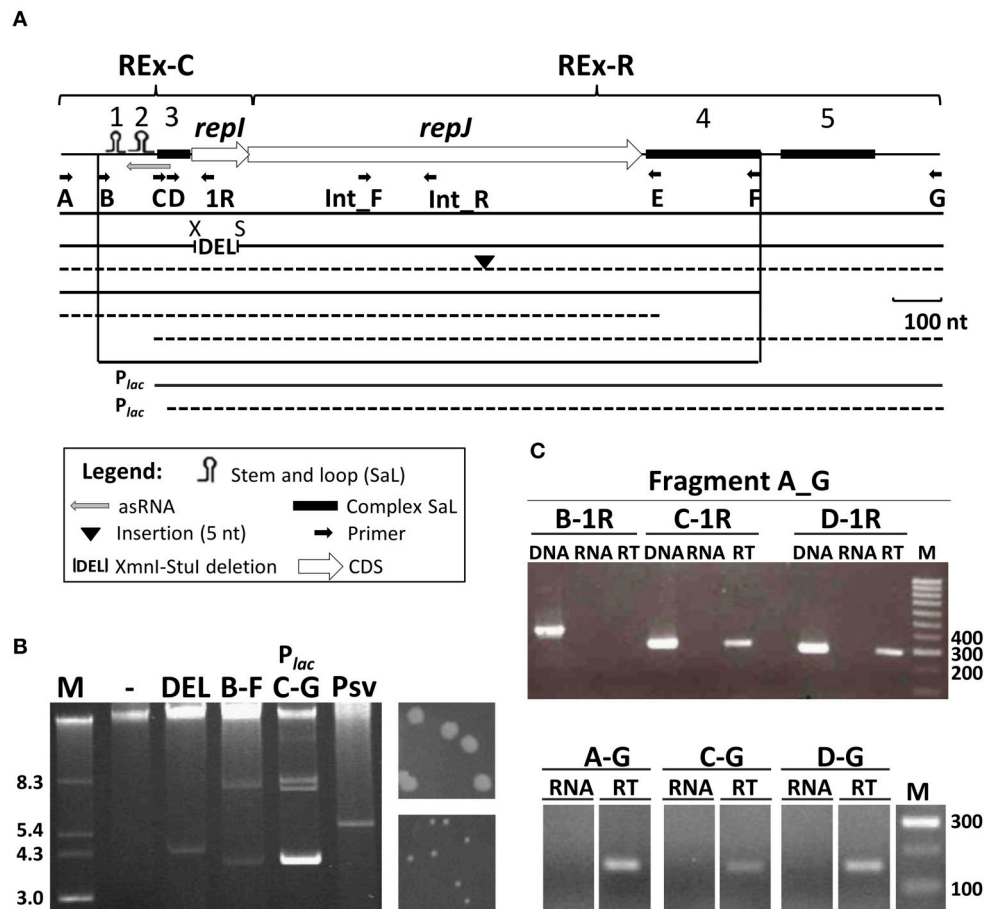
## RESULTS

### Definition of a Second Replicon in Plasmid pPsv48C

In an independent study (M. Añorga, unpublished data), we observed the spontaneous generation of autonomously replicating deletion derivatives of pPsv48C lacking the RepA-PFP replicon (**Figure 1B**). The smallest derivative contains five putative coding sequences (CDSs), whose annotation is not related to plasmid replication, and a 661 nt fragment that appears in two nearly identical copies in pPsv48C (**Figure S1**). The second copy of this fragment precedes gene *repA*, matching the typical

organization of RepA-PFP replicons, and was previously shown to be essential for replication of PFP plasmids (Murillo and Keen, 1994; Sesma et al., 1998; Gibbon et al., 1999; Sundin et al., 2004).

By cloning diverse PCR fragments into the *E. coli* vectors pK184 or pKMAG (**Figure S2**), which do not replicate in *Pseudomonas*, we determined that a 1,375 nt fragment (coordinates 29,386–30,760 in the pPsv48C sequence, accession no. FR820587; fragment B-F, **Figure 1A** and **Figure S1**), comprising around half of the 661 nt repeated fragment, contained all the essential elements for autonomous replication in the plasmidless strains *P. syringae* pv. *syringae* B728a (**Figure 1B**) and *P. syringae* pv. *savastanoi* UPN912. This



**FIGURE 1 | Functional characterization of the RepJ replicon. (A)** The minimal replicon consists of two modules, REx-C and REx-R, whose extent is loosely defined here; stem-and-loop structures 1 to 5 are indicated by symbols (see legend) topped with the structure number; letters under small arrows indicate the primers used for amplification of fragments and for RT-PCR, and a gray arrow, the putative antisense RNA. Lines under the map indicate the relative size of fragments cloned after a transcriptional terminator in pKMAG, or downstream of the *P<sub>lac</sub>* promoter in pK184. Continuous lines indicate constructs able to replicate autonomously in *P. syringae* pv. *syringae* B728a, and dashed lines indicate non-replicative fragments. X, XmnI; S, StuI. **(B)** Plasmid profiles of B728a transformants harboring constructs from panel A (indicated by letters) able to replicate and that define the minimal replicon; multiple bands correspond to different topological forms and/or multimers with the lowest bands corresponding to monomers and the top intense band to chromosomal DNA and high-molecular-weight plasmid multimers. Lane (-) is the wild type strain B728a and Psv is a derivative of NCPPB 3335 containing a 5.6 kb spontaneous deletion derivative of pPsv48C spanning the RepJ replicon. To the right, 48 h colonies of transformants containing constructs A-G (top) or C-G (bottom). **(C)** *repJ* is transcribed from at least two promoters. Upper gel: for RT-PCR, cDNA was synthesized from UPN912 containing construct A-G (in pKMAG) using the strand-specific primer Int\_R, internal to *repJ*. Subsequent PCRs were done with primer pairs indicated above the gel and shown on (A). Lower gel: for RT-PCR, cDNA molecules were synthesized from UPN912 containing constructs A-G, C-G, and D-G (the last two cloned in pKMAG-C) using the strand-specific primer Int\_R. Further PCR amplification of cDNA was done using primer pair Int\_F-Int\_R, internal to *repJ* (see panel A). Amplifications were done with only DNA (DNA), only RNA (RNA) or cDNA (RT); M, molecular weight marker in kb (panel B) or nt (panel C).

fragment did not contain any obvious direct repetitions reminiscent of iterons, but was rich in palindromic structures and could adopt a complex folding structure. We could distinguish two well-defined structural regions in this minimal replicating fragment; based on their conservation and functionality (see below), we have defined these regions as plasmid replicon *exchangeable* (REx) modules: the REx-C module contains the putative replication control system, whereas the REx-R module comprises the replication system.

### REx-C Module

This is a 318 nt fragment (coordinates 29,386–29,703 in FR820587) that shows high identity to a fragment (coordinates 41,791–5) preceding the *repA* gene from pPsv48C and including its first two codons (**Figure 1** and Figure S1). The fragment contributes the putative start codon, the promoter(s) and the RBS for the expression of the replication initiator gene *repJ* (see below). It also contains three SaL structures, the third of which is complex, potentially folding in different ways, and gene *repI*. By analogy with replicons lacking iterons (del Solar and Espinosa, 2000; Brantl, 2014), such as ColE2, this fragment also probably codes for a small antisense RNA, with a putative promoter within SaL 3 and for which SaL 1 could function as a transcription terminator (Figure S1). Replication assays with clones spanning partial fragments of the minimal replicon showed that deletion of SaL 1 and 2 abolished autonomous replication, although they were dispensable in clones maintaining the complex SaL 3 and the strong *Plac* promoter of the vector in the same transcriptional direction as *repJ* (fragment CG in **Figure 1A**). Nevertheless, bacteria transformed with construct CG required double the time than other replicative fragments to produce visible colonies (**Figure 1B**). Additionally, clones lacking the 5' stem of SaL 3 did not sustain autonomous replication (fragment DG, **Figure 1A** and Figure S1), even when *repJ* was cloned in the transcriptional direction of the *Plac* promoter. These results likely suggest that expression of *repJ* from the *Plac* in this clone causes a lethal runaway replication phenotype (Nordström and Wagner, 1994) or that SaL 3 is also essential for replication.

Gene *repI* (PSPSV\_C0037) is short (123 nt) and appears to be translationally coupled to the replication initiator gene *repJ*, which are characteristics of leader peptide genes needed for the control of replication of certain replicons (del Solar et al., 1998; Brantl, 2014). An XmnI-StuI in-frame deletion of 87 nt, spanning most of *repI* (**Figure 1A** and Figure S1) did not have any apparent effect in the replication ability of the RepJ replicon, indicating that the product of *repI* is not essential for replication and that spacing between the SaL structures and the start of gene *repJ* is flexible.

### REx-R Module

This module contains the replication initiator gene, *repJ* (PSPSV\_C0038, 819 nt) and essential downstream sequences. The long, near-perfect ribosome binding site (5'-AAGGcGGTGA-3') of *repJ* and its two first codons probably belong to the REx-C module, because they are part of a sequence highly conserved in the pPsv48C RepA-PFP replicon (Figure S1). Gene *repJ* is annotated as a putative transcriptional regulator

and did not show significant homology to any domain in an InterPro search. However, the structure of 55 residues (residues 90–145) from RepJ could be modeled by Phyre2 with 81.6 % confidence, being similar to the N-terminal domain of a conserved replication initiator protein (Schumacher et al., 2014). Additionally, a construct containing a mutation causing a premature stop in *repJ* did not replicate in the plasmidless strains B728a and UPN912 (**Figure 1**). These results suggest that *repJ* codes for a replication initiator protein essential for autonomous replication.

After the *repJ* stop codon there is a *ca.* 0.5 kb fragment containing two blocks of repeated sequences that can form complex SaL structures, designated SaL 4 and 5, although only SaL 4 appears to be essential for autonomous replication (**Figure 1A**). Nevertheless, a blastn search with this fragment identifies sequences similar to SaL 5 situated 3' of, among others, *rep* genes that are not homologous to *repJ*, such as those from plasmids pRA2 (from *P. alcaligenes* RA2), pP27494\_2 (from *P. antarctica* PAMC 27494), pMBUI6 (from an uncultured bacterium), and pAOVO01 (from *Acidovorax* sp. JS42), and gene *krfA* from plasmid pTer331 (from *Colimonas fungivorans* Ter331).

## The REx-C Module Contains at Least Two Active Promoters

The absolute requirement of SaL 1 and 2 for replication can be overcome when *repJ* is transcribed from the strong *Plac* promoter, suggesting a plausible role of these structures in directing *repJ* transcription. We thus examined transcription of this gene by RT-PCR in clones lacking SaL 1, 2 and 3. Using clone AG in pKMAG (**Figure 1C**, upper gel), we observed a long *repJ* transcript that extended at least to the annealing site for primer C, but not to that for primer B. This indicates transcription from a promoter situated between the annealing sites for primers A and C (Figure S1), overlapping the putative antisense RNA, and possibly involved in transcription of *repI* and *repJ*. Amplification from smaller clones showed that *repJ* was transcribed even in the absence of SaL 1, 2 and 3 (fragments CG and DG, **Figure 1C** lower gel). Since the vectors used contain a T4 transcriptional terminator upstream of the cloned fragments, this shows that there is an additional active promoter immediately upstream of *repJ*. These results indicate that the REx-C module contains at least two functional promoters for the transcription of gene *repJ* and that the failure of fragments CG and DG to replicate is not due to a lack of transcription of *repJ*, suggesting an additional role for SaL 1 and 2.

## The Structure of the RepJ Replicon Is Only Partially Conserved in *Pseudomonas*

Blast comparisons revealed that a region of up to 2,785 nt containing the minimal RepJ replicon (coordinates 28,783–31,067 in FR820587) is syntenic, with very high identity, in diverse genomes of the *P. syringae* group as well as, with less identity, in a few other pseudomonads (not shown). Most of the homologs are from draft genomes and it is not possible to clearly determine if they localize to the chromosome or to

plasmid sequences. Sequence variation among the homologs from *P. syringae* was not distributed randomly (Figure S3); whereas the stems from SaL 1, 2, and 3 were identical in all sequences, there was a high sequence variation in the loop of SaL 2, in the 3' end of gene *repJ* and in SaL 4 and 5, downstream of this gene. As it occurs with the *repA* gene (Gibbon et al., 1999), the nucleotide variation in *repJ* (Figure S3) leads to a higher degree of variation in the C-terminal end of the deduced product; the phylogeny of a selection of these products is shown as clade I in **Figure 2**.

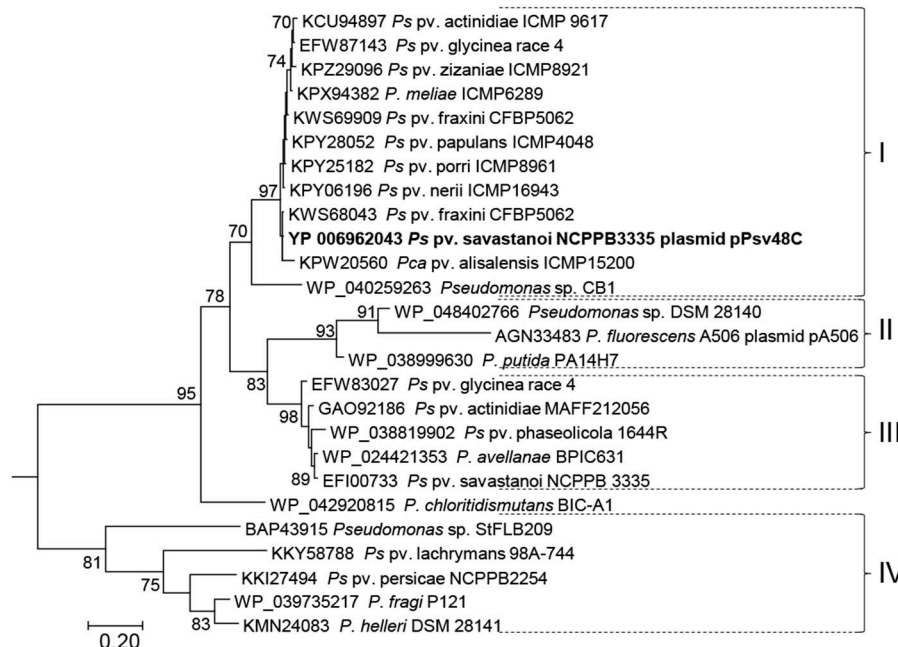
Using blastp, we found RepJ homologs only within members of Pseudomonadales, mainly in species of *Pseudomonas*, and in a few *Desulfovibrio* spp. strains. An ML tree with selected sequences (**Figure 2**) grouped homologs in four well-defined clades; from this tree we can also infer that the RepJ replicon has recently moved horizontally among pathovars and species of the *P. syringae* complex and that certain strains contained two *repJ* homologs (e.g., *P. syringae* pv. fraxini CFBP5062 in clade I). The minimal RepJ replicon is moderately conserved and syntenic among homologs from clades I and II (Figure S4). Of note, the REx-C modules from the RepJ replicons from clade I were more similar among themselves than to the modules from RepA-PFP replicons, indicating that these RepJ replicons are long-time inhabitants of pseudomonads. Conversely, the REx-C module is not conserved among members in clades III and IV whereas the region downstream of *repJ* shows a degree of conservation (Figure S4); additionally these homologs

show a genetic organization that is conserved among clade members, but different than that from members of clades I and II.

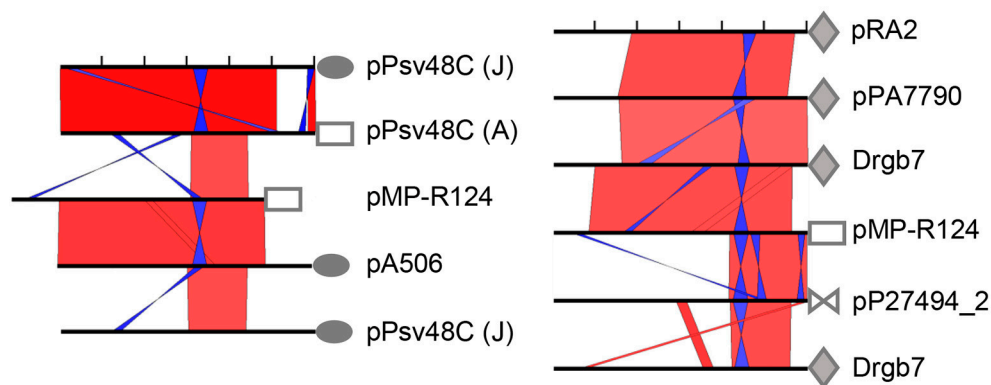
A 2,808 nt fragment (positions 35,079–37,886, accession no. KB644113) containing the *repJ* homolog from the chromosome of strain NCPPB 3335 (PSA3335\_1080, clade III; see Figure S4) cloned in pKMAG did not generate any transformant after electroporation into *P. syringae* pv. *syringae* B728a or *P. syringae* pv. *savastanoi* UPN912, whereas in the same experiments we obtained hundreds of clones using the minimal RepJ replicon from pPsv48C. These results suggest that the chromosomal homolog from strain NCPPB 3335 is not able to sustain autonomous replication.

## The REx-C Module Associates to REx-R Modules of Diverse Families and Is Exchanged among Them

Blastn comparisons showed that the REx-C module is present, with varying degrees of conservation, preceding the *rep* genes from at least four non-homologous replicon families from *Pseudomonas* (**Figure 3**; Table S1), whereas nucleotide identity is rapidly lost shortly after the CDS start codon. These families include the RepA-PFP family from *P. syringae* and other bacteria (exemplified by *repA* from pPsv48C) (Bardaji et al., 2011), the RepJ family (among others, *repJ* from pPsv48C and pA506) (Bardaji et al., 2011; Stockwell et al., 2013), and what we



**FIGURE 2 | Maximum likelihood phylogenetic tree of RepJ.** Protein sequences from species of *Pseudomonas* are indicated by their accession no. followed by strain designation; the first 39 positions of the alignment were discarded to eliminate biases due to differential annotation of the start sites. The tree was rooted with sequence WP\_052264087.1 from *Azotobacter chroococcum* plasmid pAcX50f and bootstrap percentages of 500 replicates higher than 70% are shown close to each node. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Brackets followed by roman numerals indicate clades discussed in the main text. Pca, *P. cannabina*; Pco, *P. coronafaciens*; Ps, *P. syringae*.



**FIGURE 3 | The REx-C module is conserved among four non-homologous replicon families and subjected to recombination among them.** Graphical view of NCBI blastn comparisons of the 600 nt region preceding the start codon of replication initiator protein genes (black bars, ticks every 100 nt), displayed with the ACT software; default algorithm parameters were used, except for a word size of 7. Red and blue bars indicate collinear and inverted regions of identity, respectively, with color intensity proportional to identity; only matches larger than 15 nt are shown. Symbols are followed by their corresponding plasmid name and are identical for *rep* genes of the same homology family. Sequence accession no. are pA506, NC\_021361; plasmid Drgb7, KT351738; pMP-R124, NZ\_CM001562; pP27494\_2, CP015602; pPA7790, CP015000; pPsv48C, FR820587; pRA2, U88088. A and J refer to *repA* and *repJ*, respectively, from pPsv48C.

designated the RepA-RA2 family (named after pRA2) (Kato and Mizobuchi, 1994) and the RepA-Pa family (named here after the *P. antarctica* PAMC 27494 plasmid pP27494\_2). The highest level of sequence conservation of the REx-C module among representatives of the four families is in a 175–194 nt fragment that spans SaL structures 1–3 (Figure S5). The stem sequences from SaL 1 (the putative transcriptional terminator for the antisense RNA), as well as the stretch of adenines in either side of them (Gibbon et al., 1999), are almost perfectly conserved and changes in one arm of the stems are usually compensated with complementary changes in the other arm (Figure S5) (Gibbon et al., 1999). The other structures are also well-conserved among the four replicon families; however, the number and position of palindromes, which could form SaL structures, are variable among replicons (Figure 3; Figure S5). Therefore, the stark conservation of the REx-C module suggests that it contains features universally essential for plasmid replication in species of *Pseudomonas*. Remarkably, we did not find REx-C sequences conserved in the control region of ColE2 replicons, which contain a replication initiator protein homologous to that from RepA-PFP replicons, or in any other plasmid outside of the genus *Pseudomonas*, indicating a diversity of REx-C modules among homologous theta replicons.

As described above, the REx-C modules accompanying genes *repA* and *repJ* from pPsv48C are nearly identical (Figure S1), although there is a gradient of conservation of this module along homologs of their corresponding gene families (Figure 3, Figures S3, S4) (Gibbon et al., 1999; Sesma et al., 2000; Stavriniades and Guttman, 2004). Additionally, the REx-C module preceding *repJ* from pA506 shows a very high degree of identity to that preceding *repA* from pMP-R124 (a PFP plasmid), whereas they are less similar to REx-C modules from replicons of their same family (Figure 3). These results show that there is frequent horizontal exchange of REx-C modules among RepA-PFP and RepJ replicons.

The results of multiple blast comparisons indicate that this exchange of REx-C modules also occurs with other replicon families. To illustrate this, we did a blast comparison against the non-redundant nucleotide collection (December 2016) of the 600 nt fragment immediately preceding the start codon for gene *repA* from the PFP plasmid pMP-R124, which comprises the REx-C module. The *rep* genes found immediately downstream of the first eight homologous sequences retrieved belong to the RepA-PFP, RepJ and RepA-RA2 families (Table S2). As before, and as an example, the REx-C module from pMP-R124 (RepA-PFP) shows a higher identity to pA506 (RepJ) than to the REx-C module from pP27494\_2 (RepA-PFP) (Figure 3; Table S2). Incidentally, pMP-R124 and plasmids from the RepA-RA2 and RepA-Pa families also have a shorter REx-C module lacking the putative leader peptide gene sequence (Figure 3 and Figure S5). Together, these results indicate that the REx-C modules from these four families evolved vertically within the family, but were also subjected to horizontal exchange between members of other replicon families. Likewise, they indicate that plasmids from these four families have similar mechanisms of initiation of replication and its control.

### The REx-C Module Is Functionally Exchangeable among Different Replicons

Our comparative analyses of extant sequences (Figure 3) suggest that REx-C modules are freely exchanged among different replicons. However, previous works postulated that compatibility of coexisting RepA-PFP replicons was due to specificity between the C-terminal part of RepA and the loop sequence of SaL 2 (Figures S3, S6), which are highly variable and could coevolve for complementarity (Murillo and Keen, 1994; Gibbon et al., 1999; Stavriniades and Guttman, 2004; Ma et al., 2007). Additionally, the same pattern of variation is seen in a comparison of RepJ replicons (Figure S3). We therefore tested this putative



specificity by swapping the respective SaL structures from the REx-C module (SaL fragment in **Figure 4**) and the *rep* fragments (partial *repI* plus REx-R, see **Figure 4**) from plasmids p1448A-B (RepA-PFP) and pPsv48C (RepA-PFP and RepJ) (**Figure 4** and **Figure S6**). These two RepA-PFP initiator proteins are 88% identical (92% similar; Table S1 and **Figure S6**), while they are not homologous to RepJ (Table S1). Likewise, the three replicons show a different loop in SaL 2 and 1–3 nt changes within SaL 3 (**Figure S6**). In spite of the differences in sequence, all the analyzed chimeras were able to sustain autonomous replication in the plasmidless strain *P. syringae* pv. *syringae* B728a (**Figure 4**), indicating a lack of specificity between the SaL structures from the REx-C module and the rest of the replicon.

To test whether or not sequence variations in SaL 1 (**Figures S1, S6**) would be specific for each Rep protein, we evaluated chimeric clones containing the SaL fragment from plasmid pPsv48A and the *rep* fragments of the RepA-PFP or RepJ replicons from pPsv48C. These two RepA-PFP proteins, from pPsv48A and pPsv48C, are 97% identical (Table S1 and **Figure S6**), but the pPsv48A REx-C module (and consequently the SaL fragment tested) is comparatively shorter and with several changes in SaL 1 and 2. In spite of the differences, both chimeras replicated autonomously in strain B728a (**Figure 4**), indicating that sequence variations in the loop sequence from SaL 1, and in the sequence upstream of this structure, do not significantly impact replication ability.

Finally, none of the four SaL fragments cloned in pKMAG generated any transformants when transferred to strains B728a or 1448A (the latter containing two plasmids that could supply the RepA initiator *in trans*), suggesting that the REx-C module cannot sustain replication by itself.

## The REx-C Module Confers Replication Ability to a Non-replicative *repJ* Homolog

The REx-C module showed no apparent specificity with the associated Rep protein and is readily exchanged among replicons (**Figures 3, 4**), serving as a portable putative replication control region. Therefore, it is feasible that REx-C modules could move within bacterial genomes, co-opt genes with the appropriate characteristics and thereby directly create new autonomous replicons. To broadly test this, we examined the effect of the SaL structures from the RepA-PFP and RepJ replicons from pPsv48C on the replication ability of gene PSA3335\_1080. This gene is located in the chromosome of *P. syringae* pv. *savastanoi* NCPPB 3335 and its deduced product shows 72.7% identity (81.2% similarity) with that from *repJ* (**Figure 2** and **Figure S6**; Table S1), but does not replicate autonomously (see above).

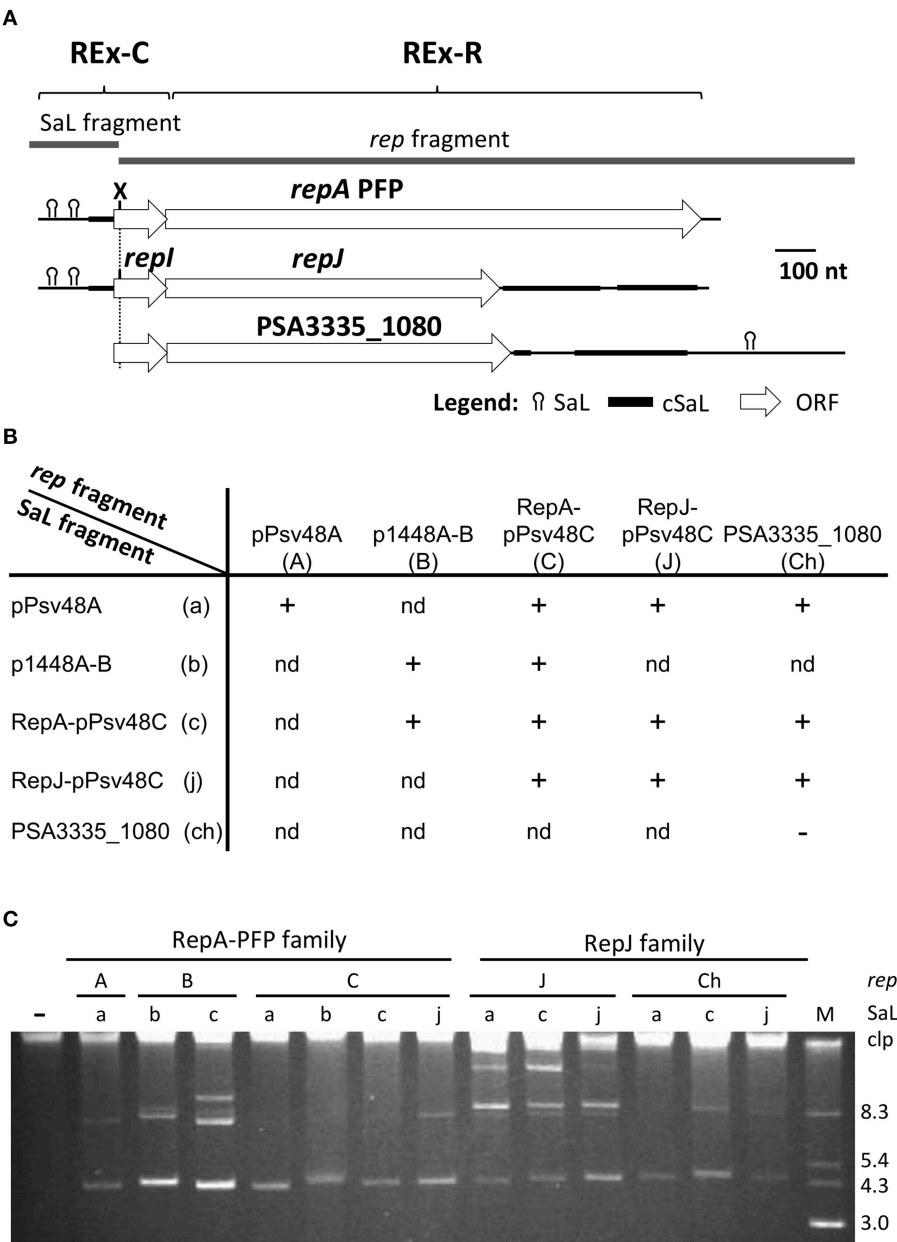
We therefore amplified and cloned a 1,772 nt fragment containing gene PSA3335\_1080 plus the upstream 114 nt, to preserve the RBS and a similar spacing to the SaL structures as with *repJ*, and 815 nt downstream, spanning sequences homologous to SaL 5 (**Figure 4**). This fragment was ligated in the proper orientation to a partial fragment of the REx-C module containing the SaL structures preceding either *repA* or *repJ* from pPsv48C (279 nt), or preceding *repA* from pPsv48A (203 nt).

The three constructions replicated in *P. syringae* pv. *syringae* B728a with very high efficiency (**Figure 4**). Disruption of the PSA3335\_1080 reading frame by filling-in an internal restriction site abolished the replication ability of the clone containing the SaL fragment from the RepA-PFP replicon of pPsv48C. These results indicate that acquisition of a REx-C module can immediately confer autonomous replication ability to sequences containing PSA3335\_1080, and that replication is dependent on the activity of this gene.

## Replicon Incompatibility Associates to the REx-C Module

Their frequent exchange (**Figure 3**), and our experiments with chimeras (**Figure 4**), indicate that the REx-C modules have a low, or no specificity for their cognate REx-R module and that they probably confer a strong selective advantage. Among other possibilities, we speculated that the exchange of these modules could be a way to reduce or evade incompatibility, especially for coexisting plasmids carrying related replicons, such as the PFP group. The Cole2 replicon, related to the RepA-PFP replicons (Gibbon et al., 1999), contains two incompatibility determinants, corresponding to the control region and to the origin of replication (Tajima et al., 1988; Hiraga et al., 1994). Importantly, the RepA-PFP replicons contain sequences highly similar to the origin of replication of Cole2 replicons (**Figure S7**) (Yagura et al., 2006), including the primer RNA sequence (AGA), and located either within the RepA coding sequences or, for pMP-R124, situated some 300 nt after the *rep* gene stop codon.

We thus evaluated possible changes in the incompatibility behavior of chimeric replicons by transforming strain *P. syringae* pv. *phaseolicola* 1448A with the native RepA-PFP replicons from plasmids p1448A-B (B) and pPsv48C (C), and with their two corresponding chimeras of SaL and *rep* fragments (combinations B-C and C-B of the SaL-*rep* fragments; see **Figures 4, 5**). Strain 1448A contains naturally the native RepA-PFP plasmids p1448A-A (132 kb) and p1448A-B (52 kb). As expected, the native replicon from pPsv48C did not show any obvious incompatibility, producing a large number of transformants and coexisting with the two native plasmids from strain 1448A (**Figure 5**, lane C-C). The same results were observed with the C-B chimera (**Figure 5**, lane C-B), containing the SaL structures from plasmid C and the *rep* fragment from plasmid B. Conversely, the native replicon from plasmid B (B-B in **Figure 5**) and the B-C chimera generated about half the number of transformants than the two other replicons tested, taking double the time to reach colonies of the same size. Incompatibility mediated by the native replicon from plasmid B and the B-C chimera was evident in plasmid profile gels, where they either cointegrated with p1448A-B or induced its loss, or appeared with an apparently reduced copy number (**Figure 5**). These results indicate that strong replicon incompatibility between RepA-PFP replicons is associated to the REx-C module and, unlike what happens with Cole2 replicons (Tajima et al., 1988; Yasueda et al., 1994), not to the Rep protein or the origin of replication and that it can be overcome by only a few nucleotide changes in this module.

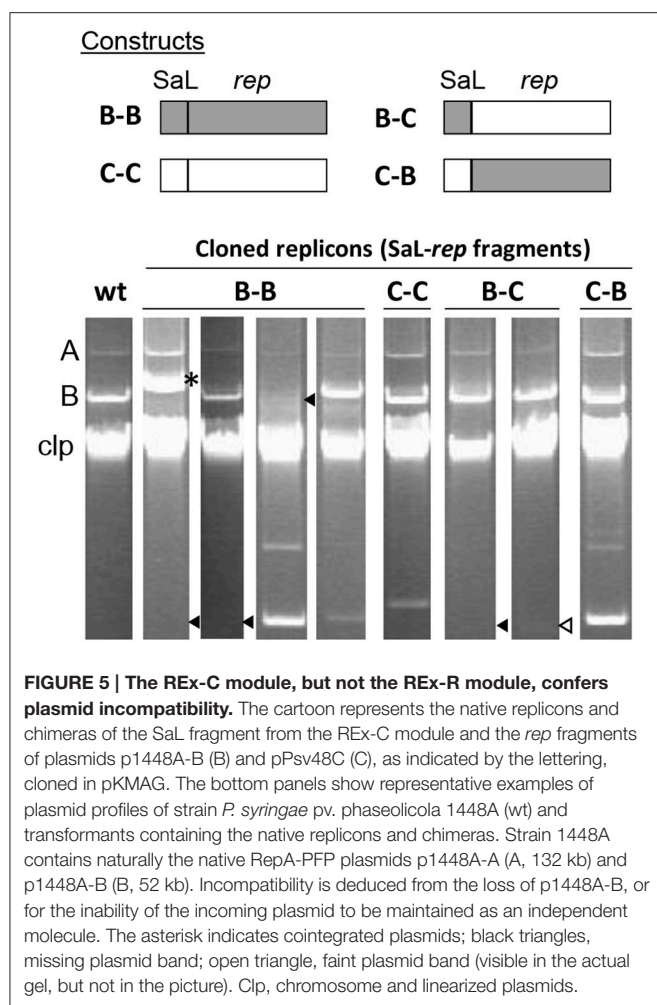


**FIGURE 4 | The REx-C modules are functionally exchangeable within and among REx-R families. (A)** Schema of the DNA fragments exchanged for the construction of chimeras, which were cloned in vector pKMAG immediately after the transcriptional terminator. X, XmnI restriction site. **(B)** Autonomous replication in *P. syringae* pv. *syringae* B728a of wild type and chimeric DNA regions: +, autonomous replication; -, no replication; nd, not determined. **(C)** Undigested plasmid profile gels of B728a harboring the different wild type and chimeric clones, with numbers indicating the combination of SaL and *rep* fragments as in **(B)**; discrete plasmid bands evidence autonomous replication, with multiple bands being different topological forms and/or multimers. Lane (-) is the wild type strain B728a, plasmidless; lane M, size markers, in kb; Clp, chromosome and high-molecular-weight plasmid multimers.

DISCUSSION

Modularity, which can be broadly defined as the degree to which a system is made up of relatively independent but interlocking parts, is ubiquitous in biology at all organization levels (Wagner et al., 2007; Kreimer et al., 2008). In this work, we demonstrate that diverse plasmid replicons from *Pseudomonas*

are also modular, being composed of discrete functional units, or modules, that are readily and frequently exchanged among unrelated systems, colliding with the traditional view of minimal replicons as heritable units that evolve as a whole and that can be classified in more or less coherent incompatibility and phylogenetic groups (del Solar et al., 1998; Petersen, 2011).



We have found extensive evidence of modularity in plasmids from many diverse species of *Pseudomonas*, including environmental species as well as plant, animal and human pathogens of paramount significance, such as *P. syringae* and *P. aeruginosa* (Figures 2, 3, Table S2, and not shown). Nevertheless, it is highly likely that this concept is also applicable to plasmid replicons consisting of *rep* gene and control systems from other organisms. In particular, homologous and/or site-specific recombination has been postulated to contribute to a similar exchange of control and *rep* modules in related enterobacterial plasmids of the IncI $\alpha$  and IncFII incompatibility groups (Kato and Mizobuchi, 1994) of the class A theta replicons, as well as different segments of ColE2 replicons (Hiraga et al., 1994); nevertheless, the concept of modularity as a general organizational model for replicons did not emerge from these studies. Based on their properties as functionally independent and exchangeable units, we defined two replicon exchangeable modules and designated them as REX-C (control) and REX-R (replication) modules, respectively (Figure 1).

Module REX-C consists of the replication control region, and has variable size and configuration. As defined for the RepJ replicon from pPsv48C, REX-C is a small sequence that

contains three potential SaL structures, a putative leader peptide, a putative antisense RNA, and the signals for transcription and translation of the *rep* gene (Figure S1). Based on these characteristics, and by structural similarity to previously described replicons, it is highly likely that the REX-C module is involved in copy number control by the antisense RNA (del Solar et al., 1998; Brantl, 2014). Sequence comparison of extant replicons (Figure 3) and functional assays (Figures 1, 4) indicate that this module is an autonomous but essential part of the replicon, and that it is functionally exchangeable among four non-homologous REX-R families (Table S1). The module contains a core region of 175–194 nt highly conserved among members of the four families (Figure 3, Figure S5) and that spans the three SaL structures essential for autonomous replication of the minimal RepJ (Figure 1) and the RepA-PFP replicons (Murillo and Keen, 1994; Gibbon et al., 1999). We showed that chimeric replicons containing different combinations of SaL structures and *rep* genes from the RepA-PFP and RepJ families were fully functional (Figure 4), demonstrating the independent functionality of the REX-C module and a lack of specificity with the REX-R module. Importantly, the *rep* fragments used for the construction of chimeras retained their corresponding peptide leader gene sequence and the RBS for the *rep* gene (Figure 4), suggesting that control of replication by the REX-C module does not involve the formation of secondary structures (pseudoknot) with sequences surrounding the *rep* gene. Additionally, the SaL structures also conferred replication ability to the *rep* fragment containing PSA3335\_1080 (Figure 4), whose sequence is poorly conserved compared with the other replicons used for the construction of chimeras. Work with pA506 suggested that clones containing its REX-C module, without the *repJ* gene, were able to sustain autonomous replication (Stockwell et al., 2013), although these authors did not conclusively eliminate the possibility of ectopic integration. Conversely, evidence presented here and elsewhere (Murillo and Keen, 1994; Kwong et al., 1998; Gibbon et al., 1999; Sundin et al., 2004) shows that diverse REX-C modules were unable to replicate autonomously. Consequently, their functionality and conservation suggest that SaL structures in the REX-C module are essential for replication and replication control of a wide variety of replicons. This likely happens because the REX-C module harbors the signals required for expression of the *rep* gene, whose accessibility and proper recognition would be modulated by the interaction between the antisense RNA and the mRNA. Since the antisense RNA and its target sequence on the *rep* mRNA would be coded for in complementary strands of the same SaL region (Figure S1), they always show a perfect complementarity independently of any other downstream sequence. Therefore, the REX-C module would then act as an independent, self-contained portable unit for the control of replication.

The SaL structures are followed by a putative leader peptide that is present in diverse replicons as part of the region for the control of plasmid copy number (del Solar et al., 1998; Brantl, 2014). Previous works (Wagner et al., 1987; Blomberg et al., 1992) also showed that the level of expression of the *rep* gene is controlled by the level of translation of the leader peptide gene, and not by its product. This is compatible with the fact that an

in-frame deletion of most of the putative leader peptide gene (*repI*) sequence in the RepJ replicon did not have any significant effect on replication (**Figure 1**). Likewise, lack of function for the product of this gene could justify the large sequence differences in the leader peptide genes of the RepJ and RepA-PFP replicons from pPsv48C, despite the rest of the REx-C module being highly similar (Figure S1). Remarkably, some of the REx-C modules examined here (e.g., from pMP-R124 and pRA2) lack the leader peptide gene, with the consequence that the start of the *rep* gene is in very close proximity to the end of SaL 3 (**Figure 3** and **Figure S5**). Nevertheless, the REx-C module is highly flexible and can accommodate different structures to ensure functionality in a diversity of replicon arrangements. For instance, functionality of the pRA2 replicon required a large region upstream of the *rep* gene, containing in this order: four potential iterons, an additional small gene (*repB*, 240 nt), and a fragment containing the conserved SaL structures shown in Figure S5 (Kwong et al., 1998).

We found four different families of REx-R modules defined by groups of homology of the corresponding Rep proteins (**Figure 3** and Table S1). The family RepA-PFP is composed of a single gene, coding for a large protein of as much as 437 amino acids homologous to many replication proteins from diverse species of Gram-negative and Gram-positive bacteria, including ColE2 replicons (Gibbon et al., 1999; Yagura et al., 2006). The RepA-PFP proteins contain highly conserved primase, PriCT and HTH domains, with the vegetative origin of replication located in the 3' end of the *rep* gene or immediately after this gene (Figure S7) (Yagura et al., 2006). The other REx-R families contain smaller *rep* genes, with deduced products ranging from 269 to 341 amino acids and lacking homology to protein families and domains included in the InterPro database (Mitchell et al., 2015). These genes might also be accompanied by downstream sequences that can form complex folding structures and that are essential for replication, as it is the case with the RepJ replicon (**Figure 1**) and pRA2 (Kwong et al., 1998). Directed mutagenesis of the *repJ* (**Figure 1**), PSA3335\_1080 and the *repA*-PFP (Gibbon et al., 1999) *rep* genes, demonstrates that they are essential for replication. Thus, the diversity of the REx-R modules found here again stresses the functional universality of the REx-C module.

The exchange of REx modules among disparate replicons is intriguing and can confer diverse evolutionary advantages, such as increasing their competitiveness by acquiring modules that are better adapted to their particular bacterial host. In this respect, the type of REx-C module characterized here is probably specific to *Pseudomonas* and its general presence in diverse replicons likely facilitates plasmid survival in these bacteria. Indeed, we were unable to find sequences homologous to the REx-C module in any other bacteria outside the genus *Pseudomonas*. Additionally, the RepA-PFP replicons analyzed here contain specific REx-C modules that are different from those associated to homologous RepA proteins, such as those from the enterobacterial ColE2 plasmids. Another likely advantage of modularity is to favor the coexistence of highly related replicons. This type of coexistence has been documented for PFP plasmids of pseudomonads (Murillo and Keen, 1994; Sundin, 2007) and *repABC*-type plasmids from rhizobia (Cevallos et al., 2002), and

it also likely happens with RepJ replicons (**Figure 2**). The analysis of chimeric replicons indicates that plasmid incompatibility is conferred by the REx-C module (**Figure 5**), in agreement with previous results indicating a partial incompatibility of the corresponding module from plasmid pPT23A (Gibbon et al., 1999). Additionally, the incompatibility between plasmid p1448A-B and its cloned replicon was bypassed by swapping the REx-C module of the cloned replicon with another one differing in the sequence of the loop from SaL 2, as well as in a few other nt positions (**Figure 5** and **Figure S6**). Therefore, the exchange of REx-C modules could be a rapid and efficient way of reducing or eliminating incompatibility among highly related PFP plasmids, which usually carry genes conferring important adaptive advantages (Vivian et al., 2001; Sundin, 2007). Additionally, it is also likely that the coexistence of related modular replicons, requiring similar cellular resources, could help to streamline the replication process and reduce the metabolic burden of plasmids.

A further and exciting consequence derived from the modularity of replicons, is that their combination could potentially produce new replicons. In particular, the independent functionality of the REx-C module could allow it to co-opt appropriate genes that might then function as Rep initiators and result in new autonomously replicating molecules. But, would the bacterial gene pool contain genes that will function as plasmid replication initiators when associated to a REx-C module? Our experiments indicate that this is indeed the case: the non-replicative chromosomal *repJ* homolog PSA3335\_1080 could replicate autonomously when preceded by a DNA fragment containing the SaL structures from the REx-C module. PSA3335\_1080 is a chromosomal gene widely distributed in *P. syringae* and related species, conserving synteny with its adjacent sequences, which suggests that is a long time inhabitant of these bacteria possibly having a functional role different from replication. A priori, there could be a variety of genes that could generate new replicons when combined with an appropriate REx-C module and their identification could be challenging. For instance, manual or automatic prediction of structure or function of the RepJ protein and homologs would likely be unsuccessful because they lack obviously conserved domains typical of Rep proteins, and this might also be the case for other genes that could be recruited as replication initiators. Therefore, and as occurs with other systems (Agapakis and Silver, 2009; Lorenz et al., 2011; Melo et al., 2016), modularity of origins of replication will undoubtedly favor their evolution and adaptability, for instance by reducing incompatibility and metabolic load to the bacterial host, but could also be facilitating the generation of new plasmid replicons, with the concomitant possibility of immediate mobilization of associated chromosomal genes.

## MATERIALS AND METHODS

### Bacterial Strains, Plasmids, and Growth Conditions

Bacterial strains and plasmids used in this work are detailed in Table S3. *E. coli* strains DH10B and GM2929 (*dcm*<sup>-</sup>, *dam*<sup>-</sup>),



when unmethylated DNA was needed, were used for DNA manipulations and were grown in LB at 37°C. Strains *P. syringae* pv. phaseolicola 1448A (Joardar et al., 2005), pv. *syringae* B728a (Feil et al., 2005), and pv. *savastanoi* NCPPB 3335 (Rodríguez-Palenzuela et al., 2010) and UPN912, which derives from strain NCPPB 3335 and is cured of its three native plasmids (M. Añorga, unpublished results), were propagated using King's medium B (King et al., 1954) at 25°C. When necessary, media were supplemented with 100 µg ml<sup>-1</sup> ampicillin or 25 µg ml<sup>-1</sup> kanamycin. We used a mixture of plasmids pME6031 (8.3 kb), pBBR1MCS-2 (5.4 kb), pKMAG-C (4.3 kb) and pBlueScript II (3.0 kb) (Table S3), purified from *E. coli* using the Illustra plasmidPrep Mini Spin kit (GE Healthcare, UK), as size markers in plasmid profile gels.

## Molecular Techniques

DNA was amplified using a high fidelity enzyme (PrimeStar HS, Takara Bio Inc., Japan) and cloned using the CloneJET PCR Cloning Kit (Thermo Scientific) or the pGEM-T Easy Vector System (Promega), following the manufacturer's instructions. For plasmid profile gels, DNA was purified by alkaline lysis and separated by electrophoresis in 0.8% agarose gels with 1xTAE as described (Murillo et al., 1994). Plasmids were transferred to *P. syringae* by electroporation (Choi et al., 2006). For RT-PCR analysis, DNA-free RNA was obtained from bacterial cultures grown overnight in medium B using TriPure Isolation Reagent (Roche Diagnostics) and Ambion TURBO DNA-free Kit (Life Technologies). Concentration and purity of RNA were determined spectrophotometrically, and its integrity confirmed by electrophoresis in agarose gels. cDNA was synthesized from RNA using the ImProm-II reverse transcriptase system (Promega), following the manufacturer's recommendations. Primer Int\_R, 5'-GCCGGTGCAGAGATACCC-3', specific for the sense transcript of *repJ*, was used for the reaction: at 25°C for 5 min for primer annealing, 60 min at 42°C for reverse transcription, and 15 min at 70°C for enzyme inactivation. cDNA was amplified using primers B (5'-CGATGTAGATTACGAATCGCAG-3'), C (5'-CTGATTATGGCGTTCACTGC-3'), or D (5'-TGCAAGCTGTCTAAAGTGAAGC-3') together with 1R (5'-GCTGTTGTTTCAGAGATGACG-3') to determine the size of the transcript, or primer pair Int\_F (5'-GAGAAGTTTC TGGCCATCGAG-3') and Int\_R (5'-GCCGGTGCAGAGATAC CC-3') to analyze the transcriptional activity of gene *repJ* (see **Figure 1**). The program used comprised 30 cycles (94°C for 30 s, 58°C for 30 s, and 72°C for 30 s) plus a final extension step of 6 min at 72°C. Control reactions included PCR amplification of pure extracted RNA, to verify the absence of contaminating DNA; amplification of purified DNA, to verify the reaction conditions, and amplification of an internal fragment of gene *gyrB*, to confirm the synthesis of cDNA.

## Bioinformatics

Multiple-sequence alignments using Muscle, determination of the optimal substitution model, and Maximum likelihood phylogenetic tree construction using the JTT matrix-based model with a gamma distribution with 5 categories were done using MEGA7 (Kumar et al., 2016); confidence levels of the branching

points were determined using 500 bootstraps replicates. Searches for sequence similarity in the NCBI databases were done using the BLAST algorithms (Hubbard et al., 2008) and sequences were aligned on-line using the MULTALIN program (Corpet, 1988) or the tools in the EMBL-EBI server (<http://www.ebi.ac.uk/Tools/msa/>). Search for protein motifs and fold recognition was done using the InterPro (Mitchell et al., 2015) (<http://www.ebi.ac.uk/interpro/>) and the Phyre2 (Kelley et al., 2015) web servers. Genome and nucleotide sequences were visualized and manipulated using the Artemis genome browser; when necessary, blast comparisons were visualized with ACT (Carver et al., 2008). Oligonucleotide primers were designed using Primer3plus software (Untergasser et al., 2012). Promoters were predicted using the online BPROM server (<http://www.softberry.com>). DNA or RNA Folding predictions were done using the Mfold web server (Zuker, 2003) using the default settings, except for a folding temperature of 25°C. Patterns of nucleotide polymorphism were calculated using DnaSP v 5.10.01 (Librado and Rozas, 2009).

## Construction of Vectors

The *E. coli* pKMAG vector and the *E. coli*-*Pseudomonas* pKMAG-C vector were constructed to examine, respectively, the replication ability of cloned fragments and the expression of *repJ* in these fragments. To construct pKMAG, a PCR-amplified fragment from the pK184 vector (positions 71-2,064; accession no. U00800) (Jobling and Holmes, 1990), retaining the kanamycin resistant gene and the origin of replication and lacking the *lacZ* promoter, and a fragment from pME6041 (Heeb et al., 2000) (positions 3,809-4,255; accession no. AF118812), containing its polylinker and the T4 transcription terminator, were ligated together resulting in the new vector. The RepA-PFP replicon from pPsv48C (positions 41,791-1,428; accession no. FR820587) was then cloned into the unique *AscI* site of pKMAG, resulting into pKMAG-C. Construction details are included in Figure S2.

## Replication Assays

Fragments used for the definition of a minimal RepJ replicon were amplified by PCR from strain Psv48ΔAB (Bardaji et al., 2011), a derivative of strain NCPPB 3335 containing only pPsv48C, cloned into pJET1.2 (Thermo Scientific) and then subcloned into pK184 in both directions, pKMAG and/or pKMAG-C. Primers used for amplifications were A (5'-AAAGCAGCGGATTTTGTAGG-3'), B, C, and D (described above), as forward, and E (5'-GACGCTAGGAGCCTATCCAG-3'), F (5'-TCCCTGTTTTCCTGAAAGG-3'), and G (5'-GGTCGAACCGACCAACTG-3'), as reverse primers (see **Figure 1** and Figure S1). For the construction of chimeric replicons we first cloned in pKMAG amplicons containing the RepA-PFP and RepJ replicons, including the REx-C module in a 390-391 nt upstream fragment, from plasmids pPsv48C (coordinates 41,714-1,336, for RepA-PFP, and 23,309-31,027, for RepJ, from accession no. FR820587) and p1448A-B (coordinates 51,624-1,650 from accession no. CP000060). The SaL (coordinates 78,056-78,262, from accession no. FR820585) and the *rep* (coordinates 78,254-1,360, from FR820585)

fragments from pPsv48A were amplified by PCR and cloned separately in pJET1.2 to reconstruct an XmnI site present in many RepA-PFP replicons and missing from pPsv48A, resulting in a T→C change in pos. 78,263. A double digestion XmnI-EcoRI or XmnI-XhoI of the resulting clones liberated the *rep* genes (*rep* fragment), leaving a 203–280 nt fragment attached to pKMAG that contained SaL structures 1 to 3 from the REx-C module (SaL fragment). The SaL and *rep* fragments were then separated by electrophoresis, purified and ligated in the appropriate combinations for the construction of chimeras. A 2,808 nt amplicon containing PSA3335\_1080 (35,079–37,886 from accession no. KB644113), including 1,150 nt upstream and 815 nt downstream of the CDS, was cloned in pKMAG to evaluate replication of this gene in its native configuration. Additionally, an amplicon similar to the previous one but lacking the first 1,036 nt (36,115–37,886 from accession no. KB644113), and with an A→T change in pos. 36,115, was ligated in the proper orientation to the appropriate XmnI-EcoRI clones in pKMAG generated above and containing the SaL fragments from *repA* or *repJ* from pPsv48C, or from pPsv48A. For functional analyses of replication initiation protein genes, replicative constructions containing *repJ* and PSA3335\_1080 were digested using a unique restriction site internal to each CDS, filled-in with Klenow enzyme (New England BioLabs Inc., UK) and subsequently religated. Enzymes used for disrupting the CDSs by changing the reading frame were SexAI, adding 5 nt to the *repJ* CDS, and EcoNI, adding 1 nt to the CDS of PSA3335\_1080. For all replication experiments, at least two amplicons from two separate amplification experiments were cloned and tested; the identity and integrity of all clones was confirmed by sequencing. Replication ability was assessed by electroporation into the plasmidless strains *P. syringae* pv. *syringae* B728a and UPN912. Plasmids were confirmed to replicate autonomously by their ability to generate antibiotic-resistant transformants and, necessarily, by their appearance as independent bands in undigested plasmid profile gels, basically as described (Murillo and Keen, 1994; Sesma et al., 1998). Experiments were repeated at least three times, with similar results.

## Incompatibility Assays

Incompatibility between native plasmids and cloned replicons was analyzed essentially as described (Nordström, 1993; Sesma et al., 1998). Briefly, constructs in pKMAG containing the native RepA-PFP replicons from plasmids pPsv48C (strain *P. syringae* pv. *savastanoi* NCPPB 3335) and p1448A-B (strain *P. syringae*

pv. *phaseolicola* 1448A) and the corresponding chimeras of SaL and *rep* fragments, were isolated from *E. coli* and individually transferred to *P. syringae* pv. *phaseolicola* 1448A, which naturally contains the native plasmids p1448A-A (132 kb) and p1448A-B (52 kb) (Joardar et al., 2005). Resulting km<sup>R</sup> transformants of strain 1448A were cultured overnight in liquid B medium with kanamycin at 25°C with shaking, and their plasmid content visualized by electrophoresis in plasmid profile gels to determine the possible eviction of native plasmids.

## ACCESSION NUMBERS

Sequences of vectors pKMAG and pKMAG-C are deposited in GenBank under accession numbers KX714576 and KX714577.

## AUTHOR CONTRIBUTIONS

LB and JM conceived the study and designed the experiments; LB and MA performed the experiments; LB, MA, JARM, GdS, and JM analyzed the data and interpreted the results; LB and JM drafted the manuscript with contributions from JARM and GdS; all authors read and approved the final manuscript.

## FUNDING

This work was funded by the Spanish Plan Nacional I+D+i grant AGL2014-53242-C2-2-R, from the Ministerio de Economía y Competitividad (MINECO), co-financed by the Fondo Europeo de Desarrollo Regional (FEDER). M.A. was supported by an FPI fellowship (reference BES-2012-054016, Ministerio de Ciencia e Innovación/Ministerio de Economía y Competitividad, Spain). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## ACKNOWLEDGMENTS

We are indebted to Pablo Llop and Theresa H. Osinga for revising the manuscript and help with English usage.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00190/full#supplementary-material>

## REFERENCES

- Agapakis, C. M., and Silver, P. A. (2009). Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Mol. Biosyst.* 5, 704–713. doi: 10.1039/b901484e
- Aviv, G., Rahav, G., and Gal-Mor, O. (2016). Horizontal transfer of the *Salmonella enterica* serovar infantis resistance and virulence plasmid pESI to the gut microbiota of warm-blooded hosts. *mBio* 7, e01395–e01316. doi: 10.1128/mBio.01395-16
- Bardaji, L., Pérez-Martínez, I., Rodríguez-Moreno, L., Rodríguez-Palenzuela, P., Sundin, G. W., Ramos, C., et al. (2011). Sequence and role in virulence of the three plasmid complement of the model tumor-inducing bacterium *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335. *PLoS ONE* 6:e25705. doi: 10.1371/journal.pone.0025705
- Blomberg, P., Nordström, K., and Wagner, E. G. (1992). Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J.* 11, 2675–2683.
- Brantl, S. (2014). Plasmid replication control by antisense RNAs. *Microbiol. Spectr.* 2, PLAS-0001-2013. doi: 10.1128/microbiolspec.plas-0001-2013
- Bruand, C., Le Chatelier, E., Ehrlich, S. D., and Jannière, L. (1993). A fourth class of theta-replicating plasmids: the pAM beta 1 family from gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11668–11672. doi: 10.1073/pnas.90.24.11668

- Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., et al. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24, 2672–2676. doi: 10.1093/bioinformatics/btn529
- Cevallos, M. A., Porta, H., Izquierdo, J., Tun-Garrido, C., Garcia-de-los-Santos, A., Dávila, G., et al. (2002). Rhizobium etli CFN42 contains at least three plasmids of the repABC family: a structural and evolutionary analysis. *Plasmid* 48, 104–116. doi: 10.1016/S0147-619X(02)00119-1
- Choi, K. H., Kumar, A., and Schweizer, H. P. (2006). A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J. Microbiol. Meth.* 64, 391–397. doi: 10.1016/j.mimet.2005.06.001
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16, 10881–10890. doi: 10.1093/nar/16.22.10881
- del Solar, G., and Espinosa, M. (2000). Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* 37, 492–500. doi: 10.1046/j.1365-2958.2000.02005.x
- del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., and Díaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 62, 434–464.
- Feil, H., Feil, W. S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11064–11069. doi: 10.1073/pnas.0504930102
- Gibbon, M. J., Sesma, A., Canal, A., Wood, J. R., Hidalgo, E., Brown, J., et al. (1999). Replication regions from plant-pathogenic *Pseudomonas syringae* plasmids are similar to ColE2-related replicons. *Microbiology* 145, 325–334. doi: 10.1099/13500872-145-2-325
- Halary, S., Leigh, J. W., Cheaib, B., Lopez, P., and Baptiste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci. U.S.A.* 107, 127–132. doi: 10.1073/pnas.0908978107
- Heeb, S., Itoh, Y., Nishijyo, T., Schneider, U., Keel, C., Wade, J., et al. (2000). Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. *Mol. Plant Microbe Interact.* 13, 232–237. doi: 10.1094/MPMI.2000.13.2.232
- Hiraga, S., Sugiyama, T., and Itoh, T. (1994). Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* 176, 7233–7243. doi: 10.1128/jb.176.23.7233-7243.1994
- Hubbard, T. J. P., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., et al. (2008). Ensembl 2009. *Nucleic Acids Res.* 37, D690–D697. doi: 10.1093/nar/gkn828
- Ito, H., Yagura, M., Shirakihara, Y., and Itoh, T. (2015). Structural basis for replication origin unwinding by an initiator primase of plasmid ColE2-P9: DUPLICATION UNWINDING BY A SINGLE PROTEIN. *J. Biol. Chem.* 290, 3601–3611. doi: 10.1074/jbc.M114.595645
- Jackson, R. W., Vinatzer, B., Arnold, D. L., Dorus, S., and Murillo, J. (2011). The influence of the accessory genome on bacterial pathogen evolution. *Mob. Genet. Elements* 1, 55–65. doi: 10.4161/mge.1.1.16432
- Joardar, V., Lindeberg, M., Jackson, R. W., Selengut, J., Dodson, R., Brinkac, L. M., et al. (2005). Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.* 187, 6488–6498. doi: 10.1128/JB.187.18.6488-6498.2005
- Jobling, M. G., and Holmes, R. K. (1990). Construction of vectors with the p15a replicon, kanamycin resistance, inducible lacZ alpha and pUC18 or pUC19 multiple cloning sites. *Nucleic Acids Res.* 18, 5315–5316. doi: 10.1093/nar/18.17.5315
- Johnson, T. A., Stedtfeld, R. D., Wang, Q., Cole, J. R., Hashsham, S. A., Looft, T., et al. (2016). Clusters of antibiotic resistance genes enriched together stay together in swine agriculture. *mBio* 7, e02214–e02215. doi: 10.1128/mBio.02214-15
- Kato, A., and Mizobuchi, K. (1994). Evolution of the replication regions of IncI alpha and IncFII plasmids by exchanging their replication control systems. *DNA Res.* 1, 201–212.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- King, E. O., Ward, N. K., and Raney, D. E. (1954). Two simple media for the demonstration of pyocyanin and fluorescein. *J. Lab. Clin. Med.* 44, 301–307.
- Kreimer, A., Borenstein, E., Gophna, U., and Rupp, E. (2008). The evolution of modularity in bacterial metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6976–6981. doi: 10.1073/pnas.0712149105
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kwong, S. M., Yeo, C. C., Chuah, D., and Poh, C. L. (1998). Sequence analysis of plasmid pRA2 from *Pseudomonas alcaligenes* NCIB 9867 (P25X) reveals a novel replication region. *FEMS Microbiol. Lett.* 158, 159–165. doi: 10.1111/j.1574-6968.1998.tb12815.x
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Lilly, J., and Camps, M. (2015). Mechanisms of theta plasmid replication. *Microbiol. Spectr.* 3, PLAS-0029-2014. doi: 10.1128/microbiolspec.plas-0029-2014
- Lorenz, D. M., Jeng, A., and Deem, M. W. (2011). The emergence of modularity in biological systems. *Phys. Life Rev.* 8, 129–160. doi: 10.1016/j.plrev.2011.02.003
- Ma, Z., Smith, J. J., Zhao, Y., Jackson, R. W., Arnold, D. L., Murillo, J., et al. (2007). Phylogenetic analysis of the pPT23A plasmid family of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 73, 1287–1295. doi: 10.1128/AEM.01923-06
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x
- Melo, D., Porto, A., Cheverud, J. M., and Marroig, G. (2016). Modularity: genes, development, and evolution. *Annu. Rev. Ecol. Syst.* 47, 463–486. doi: 10.1146/annurev-ecolsys-121415-032409
- Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43, D213–D221. doi: 10.1093/nar/gku1243
- Murillo, J., and Keen, N. T. (1994). Two native plasmids of *Pseudomonas syringae* pathovar tomato strain PT23 share a large amount of repeated DNA, including replication sequences. *Mol. Microbiol.* 12, 941–950. doi: 10.1111/j.1365-2958.1994.tb01082.x
- Murillo, J., Shen, H., Gerhold, D., Sharma, A., Cooksey, D. A., and Keen, N. T. (1994). Characterization of pPT23B, the plasmid involved in syringolide production by *Pseudomonas syringae* pv. *tomato* PT23. *Plasmid* 31, 275–287. doi: 10.1006/plas.1994.1029
- Nordström, K. (1993). “Plasmid replication and maintenance,” in *Plasmids: A Practical Approach*, 2nd Edn., ed K. G. Hardy (New York, NY: Oxford University Press), 1–38.
- Nordström, K., and Wagner, E. G. (1994). Kinetic aspects of control of plasmid replication by antisense RNA. *Trends Biochem. Sci.* 19, 294–300. doi: 10.1016/0968-0004(94)90008-6
- Novick, R. P. (1987). Plasmid incompatibility. *Microbiol. Rev.* 51, 381–395.
- Petersen, J. (2011). Phylogeny and compatibility: plasmid classification in the genomics era. *Arch. Microbiol.* 193, 313–321. doi: 10.1007/s00203-011-0686-9
- Ramos, J. L. (ed.). (2004). *Pseudomonas: Volume 1 Genomics, Life Style and Molecular Architecture*. New York, NY: Kluwer Academic-Plenum Publishers.
- Rodríguez-Palenzuela, P., Matas, I. M., Murillo, J., López-Solanilla, E., Bardaji, L., Pérez-Martínez, I., et al. (2010). Annotation and overview of the *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335 draft genome reveals the virulence gene complement of a tumour-inducing pathogen of woody hosts. *Environ. Microbiol.* 12, 1604–1620. doi: 10.1111/j.1462-2920.2010.02207.x
- Romanchuk, A., Jones, C. D., Karkare, K., Moore, A., Smith, B. A., Jones, C., et al. (2014). Bigger is not always better: transmission and fitness burden of ~1 MB *Pseudomonas syringae* megaplasmid pMP1a107. *Plasmid* 73, 16–25. doi: 10.1016/j.plasmid.2014.04.002
- Ruiz-Masó, J. A., Machón, C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M., and Del Solar, G. (2015). Plasmid rolling-circle replication. *Microbiol. Spectr.* 3, PLAS-0035-2014. doi: 10.1128/microbiolspec.plas-0035-2014
- Schumacher, M. A., Tonthat, N. K., Kwong, S. M., Chinnam, N. B., Liu, M. A., Skurray, R. A., et al. (2014). Mechanism of staphylococcal multidrug resistance plasmid replication origin assembly by the RepA protein. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9121–9126. doi: 10.1073/pnas.1406065111

- Sesma, A., Sundin, G. W., and Murillo, J. (1998). Closely related replicons coexisting in the phytopathogen *Pseudomonas syringae* show a mosaic organization of the replication region and altered incompatibility behavior. *Appl. Environ. Microbiol.* 64, 3948–3953.
- Sesma, A., Sundin, G. W., and Murillo, J. (2000). Phylogeny of the replication regions of pPT23A-like plasmids from *Pseudomonas syringae*. *Microbiology* 146, 2375–2384. doi: 10.1099/00221287-146-10-2375
- Stavrinides, J., and Guttman, D. (2004). Nucleotide sequence and evolution of the five-plasmid complement of the phytopathogen *Pseudomonas syringae* pv. *maculicola* ES4326. *J. Bacteriol.* 186, 5101–5115. doi: 10.1128/JB.186.15.5101-5115.2004
- Stockwell, V. O., Davis, E. W. II, Carey, A., Shaffer, B. T., Mavrodi, D. V., Hassan, K. A., et al. (2013). pA506, a conjugative plasmid of the plant epiphyte *Pseudomonas fluorescens* A506. *Appl. Environ. Microbiol.* 79, 5272–5282. doi: 10.1128/AEM.01354-13
- Summers, D. K. (1996). *The Biology of Plasmids*. Oxford; Cambridge, MA: Blackwell Science. doi: 10.1002/9781444313741
- Sundin, G. W. (2007). Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. *Annu. Rev. Phytopathol.* 45, 129–151. doi: 10.1146/annurev.phyto.45.062806.094317
- Sundin, G. W., Mayfield, C. T., Zhao, Y., Gunasekera, T. S., Foster, G. L., and Ullrich, M. S. (2004). Complete nucleotide sequence and analysis of pPSR1 (72,601 bp), a pPT23A-family plasmid from *Pseudomonas syringae* pv. *syringae* A2. *Mol. Genet. Genomics* 270, 462–476. doi: 10.1007/s00438-003-0945-9
- Sundin, G. W., and Murillo, J. (2009). “Gene traders: characteristics of native plasmids from plant pathogenic bacteria,” in *Plant Pathogenic Bacteria: Genomics and Molecular Biology*, ed R. W. Jackson (Cambs: Caister Academic Press), 295–310.
- Tajima, Y., Horii, T., and Itoh, T. (1988). Replication of ColE2 and ColE3 plasmids: two ColE2 incompatibility functions. *Mol. Gen. Genet.* 214, 451–455. doi: 10.1007/BF00330479
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596
- Vivian, A., Murillo, J., and Jackson, R. W. (2001). The role of plasmids in phytopathogenic bacteria: mobile arsenals? *Microbiology* 147, 763–780. doi: 10.1099/00221287-147-4-763
- Wagner, E. G., von Heijne, J., and Nordström, K. (1987). Control of replication of plasmid R1: translation of the 7k reading frame in the RepA mRNA leader region counteracts the interaction between CopA RNA and CopT RNA. *EMBO J.* 6, 515–522.
- Wagner, G. P., Pavlicev, M., and Cheverud, J. M. (2007). The road to modularity. *Nat. Rev. Genet.* 8, 921–931. doi: 10.1038/nrg2267
- Yagura, M., Nishio, S. Y., Kurozumi, H., Wang, C. F., and Itoh, T. (2006). Anatomy of the replication origin of plasmid ColE2-P9. *J. Bacteriol.* 188, 999–1010. doi: 10.1128/JB.188.3.999-1010.2006
- Yasueda, H., Takechi, S., Sugiyama, T., and Itoh, T. (1994). Control of ColE2 plasmid replication: negative regulation of the expression of the plasmid-specified initiator protein, Rep, at a posttranscriptional step. *Mol. Gen. Genet.* 244, 41–48. doi: 10.1007/BF00280185
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bardaji, Añorga, Ruiz-Masó, del Solar and Murillo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Small, Enigmatic Plasmids of the Nosocomial Pathogen, *Acinetobacter baumannii*: Good, Bad, Who Knows?

Soo Sum Lean<sup>1</sup> and Chew Chieng Yeo<sup>2\*</sup>

<sup>1</sup> Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, <sup>2</sup> Faculty of Medicine, Biomedical Research Centre, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

## OPEN ACCESS

### Edited by:

Feng Gao,  
Tianjin University, China

### Reviewed by:

Christopher Morton Thomas,  
University of Birmingham,  
United Kingdom  
Jordi Vila Estapé,  
Hospital Clinic of Barcelona, Spain

### \*Correspondence:

Chew Chieng Yeo  
chewchieng@gmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 11 May 2017

**Accepted:** 31 July 2017

**Published:** 15 August 2017

### Citation:

Lean SS and Yeo CC (2017) Small, Enigmatic Plasmids of the Nosocomial Pathogen, *Acinetobacter baumannii*: Good, Bad, Who Knows?. *Front. Microbiol.* 8:1547. doi: 10.3389/fmicb.2017.01547

*Acinetobacter baumannii* is a Gram-negative nosocomial pathogen that has become a serious healthcare concern within a span of two decades due to its ability to rapidly acquire resistance to all classes of antimicrobial compounds. One of the key features of the *A. baumannii* genome is an open pan genome with a plethora of plasmids, transposons, integrons, and genomic islands, all of which play important roles in the evolution and success of this clinical pathogen, particularly in the acquisition of multidrug resistance determinants. An interesting genetic feature seen in majority of *A. baumannii* genomes analyzed is the presence of small plasmids that usually ranged from 2 to 10 kb in size, some of which harbor antibiotic resistance genes and homologs of plasmid mobilization genes. These plasmids are often overlooked when compared to their larger, conjugative counterparts that harbor multiple antibiotic resistance genes and transposable elements. In this mini-review, we will examine our current knowledge of these small *A. baumannii* plasmids and look into their genetic diversity and phylogenetic relationships. Some of these plasmids, such as the Rep-3 superfamily group and the pRAY-type, which has no recognizable replicase genes, are quite widespread among diverse *A. baumannii* clinical isolates worldwide, hinting at their usefulness to the lifestyle of this pathogen. Other small plasmids especially those from the Rep-1 superfamily are truly enigmatic, encoding only hypothetical proteins of unknown function, leading to the question of whether these small plasmids are “good” or “bad” to their host *A. baumannii*.

**Keywords:** *Acinetobacter baumannii*, small plasmids, antibiotic resistance genes, mobilizable plasmids, Rep-1 superfamily, Rep-3 superfamily, pRAY plasmids, toxin-antitoxin

## INTRODUCTION

*Acinetobacter baumannii* is a Gram-negative nosocomial pathogen that has become a serious healthcare concern especially in the last two decades due to its rapid ability to acquire antimicrobial resistance leading to the development of pandrug resistant (PDR) isolates that are resistant to all classes of antimicrobial compounds (Magiorakos et al., 2012; Göttig et al., 2014; Lean et al., 2014). Advances in genome sequencing and their increasing affordability have led to the availability of a plethora of *A. baumannii* genomes in the public databases (Peleg et al., 2012; Liu et al., 2013; Lean et al., 2015, 2016; Wallace et al., 2016). One of the key features of the *A. baumannii* genome

is an open pan genome with a wide variety of mobile genetic elements, particularly integrons and transposons in genomic islands, some of which are known as resistance islands due to the presence of multiple antibiotic resistance genes (Fournier et al., 2006; Bonnin et al., 2012; Ramírez et al., 2013). Resistance genes are also plasmid-borne and in *A. baumannii*, plasmids range from as small as 2 kb to more than 100 kb in size (Gallagher et al., 2015; Hamidian et al., 2016a,b). The large plasmids of *A. baumannii* are often the focus of analyses due mainly to the presence of multiple antibiotic resistance genes and the self-transmissible nature of these plasmids (Hamidian et al., 2014a,b, 2016a; Hamidian and Hall, 2014) although small plasmids have been highlighted especially those that harbor antibiotic resistance genes (D'Andrea et al., 2009; Merino et al., 2010; Grosso et al., 2012; Hamidian et al., 2012, 2016b). Despite the importance of plasmids in the potential transmission of resistance and virulence genes in *A. baumannii*, there has been surprisingly very little experimental work done on the basic biology of these plasmids. We know next-to-nothing with regards to the basic replicons of these plasmids, their replication mechanisms and transmissibility. The rapidly increasing volume of *Acinetobacter* plasmid sequences in the databases from numerous whole genome sequencing projects has led to often conflicting and chaotic annotations, complicating their *in silico* analyses, a fact that was recently highlighted for all plasmid sequences in an excellent review paper by Thomas et al. (2017). So far, *A. baumannii* plasmids have been classified according to their replicase (Rep) proteins with Bertini et al. (2010) showing that there are 19 homology groups (GR1–GR19) and developing a plasmid-based replicon typing scheme based on their *rep* genes. In this mini-review, we shall examine our current knowledge of the small plasmids of *A. baumannii* (for this purpose, we shall define “small” as any plasmid that is around 10 kb and less) and present their genetic diversity and phylogenetic relationships. We will also discuss the importance of these small plasmids to their host *A. baumannii*.

## THE REP-3 SUPERFAMILY PLASMIDS

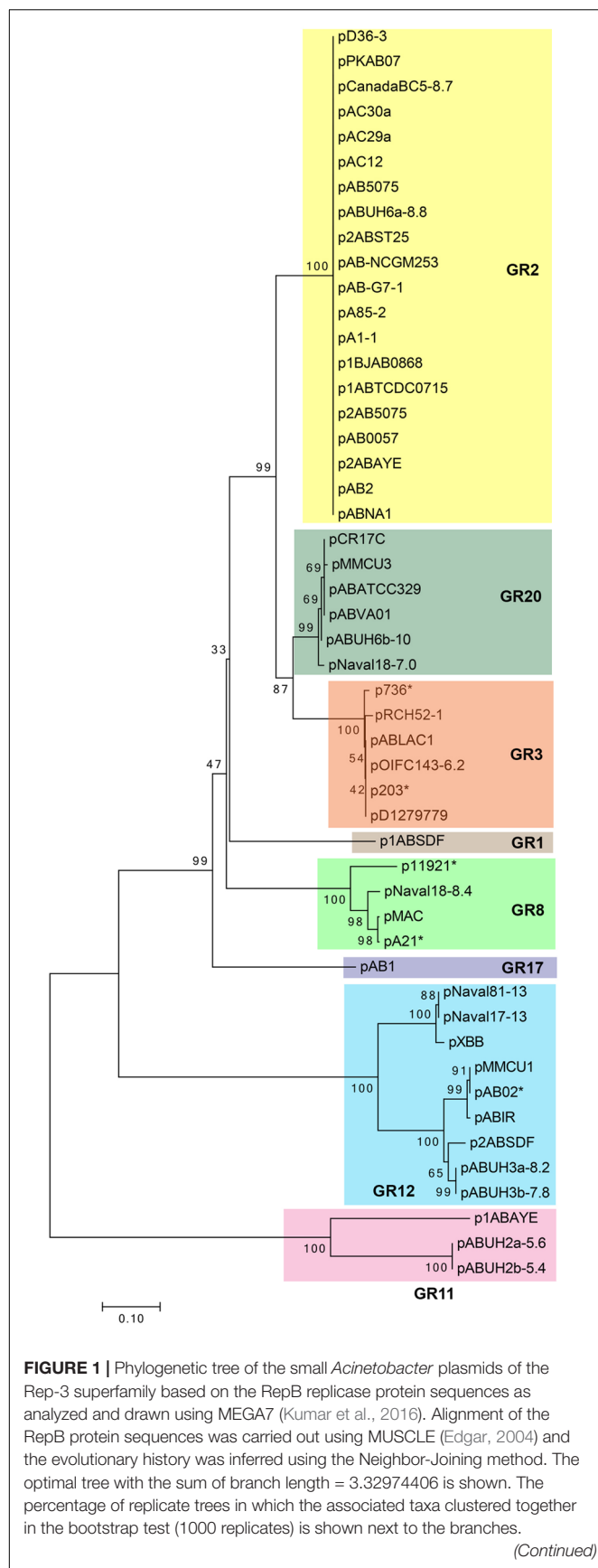
Majority of plasmids from *A. baumannii* encode replicase proteins belonging to the Rep-3 superfamily (identified by the pfam0151 conserved domain) with the larger plasmids usually harboring more than one replicon type (Bertini et al., 2010). In most of the Rep-3 superfamily replicons, the *rep* gene, which is usually annotated as *repB*, is preceded by three to six direct repeats (19–22 nucleotides in length and mainly located between 10 and 200 bp upstream of the *repB* start codon; majority are four direct repeats) that could be considered as the iterons for the RepB basic replicon (please see **Supplementary Table S1** and Data Sheet 1 for further details). In enterobacterial plasmids, these iterons serve as the origin of replication whereby the replication initiation protein binds and interacts with other host proteins (such as DnaA and the DnaBC helicase complex) required for replication initiation (Bertini et al., 2010; Konieczny et al., 2014). To the best of our knowledge, there has only been one experimental demonstration of the functionality of the *Acinetobacter* basic replicon. Dorsey et al. (2006) showed

that the minimal replicon for the 9,540-bp plasmid pMAC from *A. baumannii* 19,606 was the *repB* gene [denoted as open reading frame-1 (ORF1)] and the four direct repeats that preceded the gene in experiments using the *Escherichia coli* cloning vector pCR-Blunt II-TOPO and *Acinetobacter calcoaceticus* BD413 as host.

Phylogenetic analysis using the RepB protein sequences of 50 of these Rep-3 superfamily plasmids (**Figure 1**) was largely in agreement with the plasmid homology groups proposed by Bertini et al. (2010). However, we are of the opinion that pABVA01 which was categorized under the GR2 group by Bertini et al. (2010) warrants a separate grouping along with similar plasmids such as pMMCU3 and pABATCC329, which we designate GR20, as the phylogenetic tree clearly showed that this group of plasmids belonged to a separate clade (**Figure 1**).

Interestingly, in a majority of these small *A. baumannii* plasmids that belonged to the Rep-3 superfamily, the reading frame immediately downstream of the *repB* gene is highly conserved and is usually annotated as “*repA*” (**Supplementary Figure S1**). We could not find any homology to known replicase proteins for the translated “*repA*” gene and we are uncertain as to why this reading frame was designated *repA* in the absence of homology and/or experimental evidence. The translated protein contains a DNA-binding helix-turn-helix motif at its N-terminus and is usually annotated as a “conserved hypothetical protein” or a “DNA-binding protein” in the various database entries. The pMAC plasmid harbors this gene, which was designated ORF2, and which was shown by RT-PCR to be actively transcribed (Dorsey et al., 2006). Although for the pMAC plasmid, ORF2 was shown not to be part of the minimal replicon (Dorsey et al., 2006), its conservation in a vast majority of the small Rep-3 superfamily plasmids is suggestive of its importance. We have not found any evidence so far of the existence of any *Acinetobacter* plasmid that harbors only this “*repA*” reading frame without the *repB* gene. Nevertheless, a small number of *repB*-only plasmids do exist (such as p1ABAYE and the pABUH2a plasmids) and they form a distinct clade in the RepB phylogenetic tree (grouped under GR11; **Figure 1**) with their own unique iteron sequences (**Supplementary Table S1** and Data Sheet 1). Hence, in the absence of further experimental evidence, we could neither confirm nor completely rule out the involvement of this “*repA*” gene in the replication function of this group of plasmids. It is possible that some of these plasmids do require two replication genes, similar to IncQ plasmids such as RSF1010 which contained three replication genes with RepA functioning as the helicase, RepB as the primase, and RepC as the iteron-binding *oriV*-activator (Meyer, 2009).

Another key feature found in majority of the small Rep-3 superfamily plasmids is XerC/XerD recombination sites flanking various gene modules (**Supplementary Figure S1** and **Table S1**). Some of these gene modules include antibiotic resistance determinants such as *bla*<sub>OXA-24</sub>/*bla*<sub>OXA-40</sub> (in pABVA01, pMMCU3, pABATCC329, pABUH3a-8.2, and pABUH2a-5.6), *bla*<sub>OXA-72</sub> (in p2ABST25 and pAB-NCGM253), and the *tet*(39) tetracycline-resistance gene (in pRCH52-1). XerC and XerD recombinases usually function to convert plasmid and chromosomal dimers to monomers during cell division with

**FIGURE 1 |** Continued

The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 50 RepB amino acid sequences with the GenBank accession numbers of the plasmids as listed in **Supplementary Table S1**. Each clade of the tree corresponded with the plasmid homology grouping (GR classification) as proposed by Bertini et al. (2010) and indicated by different colored boxes. Plasmid names marked with an asterisk (\*) indicate partial plasmid sequences that covered only the *oriV-repB* sequences and were included in the analysis to validate the plasmid groupings as they were used by Bertini et al. (2010) in their classification scheme.

each recombinase catalyzing the exchange of a specific pair of strands between the recombining sites via a Holliday Junction, which is an essential reaction intermediate (Midonet and Barre, 2014). These recombinases are also involved in the integration of phage CTX- $\Phi$  in the *Vibrio cholerae* genome (Val et al., 2005) and transposition of certain conjugative transposons (Bui et al., 2006; Midonet and Barre, 2014). The DNA sequence of these small plasmids strongly infer the involvement of the XerC/XerD recombination system in the mobilization of discrete DNA modules, including antibiotic resistance genes, in *A. baumannii* although experimental proof of this has yet to be demonstrated.

Type II toxin-antitoxin (TA) systems are also found in most of the Rep-3 superfamily group of small plasmids. Type II toxin-antitoxin systems are known to mediate the stable maintenance of plasmids which harbor them through the post-segregational killing of any plasmid-free daughter cells that developed, making it difficult for the host cells to lose these plasmids (Hayes, 2003). Their presence may partly explain the widespread prevalence of this group of plasmids among *A. baumannii*. The AbkB/AbkA TA system (also known as SplT/SplA) has been shown to be a functional TA system with the AbkB (or SplT) toxin as an endoribonuclease and translational inhibitor, and AbkA (or SplA) as its cognate antitoxin (Jurenaite et al., 2013; Mosqueda et al., 2014). Other TA pairs found in these plasmids in place of AbkB/AbkA include RelE/Cro-CI (in pMAC, pABLAC1, and pD36-3), *phd-yoeB* (in p1ABAYE), *dinJ-yafQ* (in pABUH2 plasmids), and *rnlA-rnlB* (found flanked by XerC/XerD sites in pNaval18-7.0) (**Supplementary Table S1**). The functionality of these putative TA systems has yet to be experimentally verified.

Some of these Rep-3 superfamily small plasmids also harbor putative virulence factors in the form of a TonB-dependent receptor, septicolysin (Lean et al., 2016) and Sell-repeat protein. TonB-dependent receptors are known to play a role in iron acquisition (Zimble et al., 2013) whereas septicolysins are thiol-activated cytotoxins with cytolytic activity toward eukaryotic cells and have been implicated in the pathogenesis of bacteria such as *Clostridium perfringens*, *Listeria monocytogenes*, and *Streptococcus pneumoniae* (Billington et al., 2000). Sell-repeat proteins have diverse biological roles, often as adaptor proteins for the assembly of macromolecular complexes (Mittl and Schneider-Brachert, 2007). Bacterial Sell-repeat proteins mediate interactions between the pathogen and its eukaryotic host cells and have been described in

*Helicobacter pylori*, *Legionella pneumophila*, and *Pseudomonas aeruginosa* as important virulence factors, as reviewed in Mittl and Schneider-Brachert (2007). In *Neisseria meningitidis*, a Sell-repeat protein, NMB0419, was shown to be involved in meningococcal interactions with epithelial cells (Li et al., 2003) and in a recent paper, it was intriguingly shown that the expression of NMB0419 led to transcriptional changes in genes involved in iron uptake, energy metabolism, and virulence functions in a manner counteracting the global regulator, Fur (Li et al., 2017). It would therefore be of interest to experimentally investigate if these genes encoded by some of the small plasmids of the Rep-3 superfamily truly function as virulence factors for *A. baumannii*, thereby contributing to the pathogenicity of the bacterium.

Some of these small Rep-3 superfamily plasmids also encode orthologs of the MobL or MobA mobilization proteins identified by the pfam03389 conserved domain found in the MobA/MobL protein family. Plasmids that encode genes for these proteins are mobilizable by other self-transmissible plasmids. Nevertheless, the only experimental evidence for the mobilization potential of these plasmids was for pMAC of *A. baumannii* 19606 with the experiment carried out using the cloned *mobA/mobL* gene in an *E. coli* DH5 $\alpha$  host and an *E. coli* HB101 recipient (Dorsey et al., 2006). Until now, the mobilization potential of this group of plasmids from an *Acinetobacter* donor to an *Acinetobacter* recipient has yet to be shown.

## THE REP-1 SUPERFAMILY

There is a group of small cryptic plasmids from *A. baumannii* that usually comprise of a single *rep* gene and between two and five hypothetical genes. The *rep* gene of this group of plasmids encodes a replicase of the Rep-1 superfamily. Phylogenetic analysis of the Rep proteins from this group of plasmids showed that they could be divided into two subgroups: the p4ABAYE subgroup and the Rep63 subgroup (Supplementary Figure S2). The 2,726 bp p4ABAYE from *A. baumannii* AYE encodes a *rep* gene and four hypothetical ORFs (Fournier et al., 2006) and was categorized under the GR14 group of *Acinetobacter* plasmids (Bertini et al., 2010). The second subgroup contained two of the smallest reported *Acinetobacter* plasmids, the 1,967 bp p3AB5075 from *A. baumannii* AB5075 (Gallagher et al., 2015) and the 1,958 bp pM131-10 plasmid from *Acinetobacter* sp. M131 (accession no. JX101639). The small size of p3AB5075 has been validated by plasmid extraction and agarose gel electrophoresis (Gallagher et al., 2015), and the plasmid consisted of the *rep* gene and two other reading frames of unknown function. Although Gallagher et al. (2015) stated that the *rep* gene of p3AB5075 was of undefined plasmid replication group, our phylogenetic analysis indicated that it is grouped with an unpublished 2,343 bp *A. baumannii* plasmid pAB49 (accession no. L77992.1) (Supplementary Figure S2), which was previously categorized by Bertini et al. (2010) under the GR16 group. Furthermore, the *rep*-encoded protein of pAB49 had been previously shown to be homologous to the Rep63 replication initiation protein encoded by pBL63.1 of *Bacillus*

*licheniformis* and orthologs in rolling-circle replicating (RCR) plasmids from various other bacterial species (Guglielmetti et al., 2005).

## PLASMID pRAY AND ITS DERIVATIVES

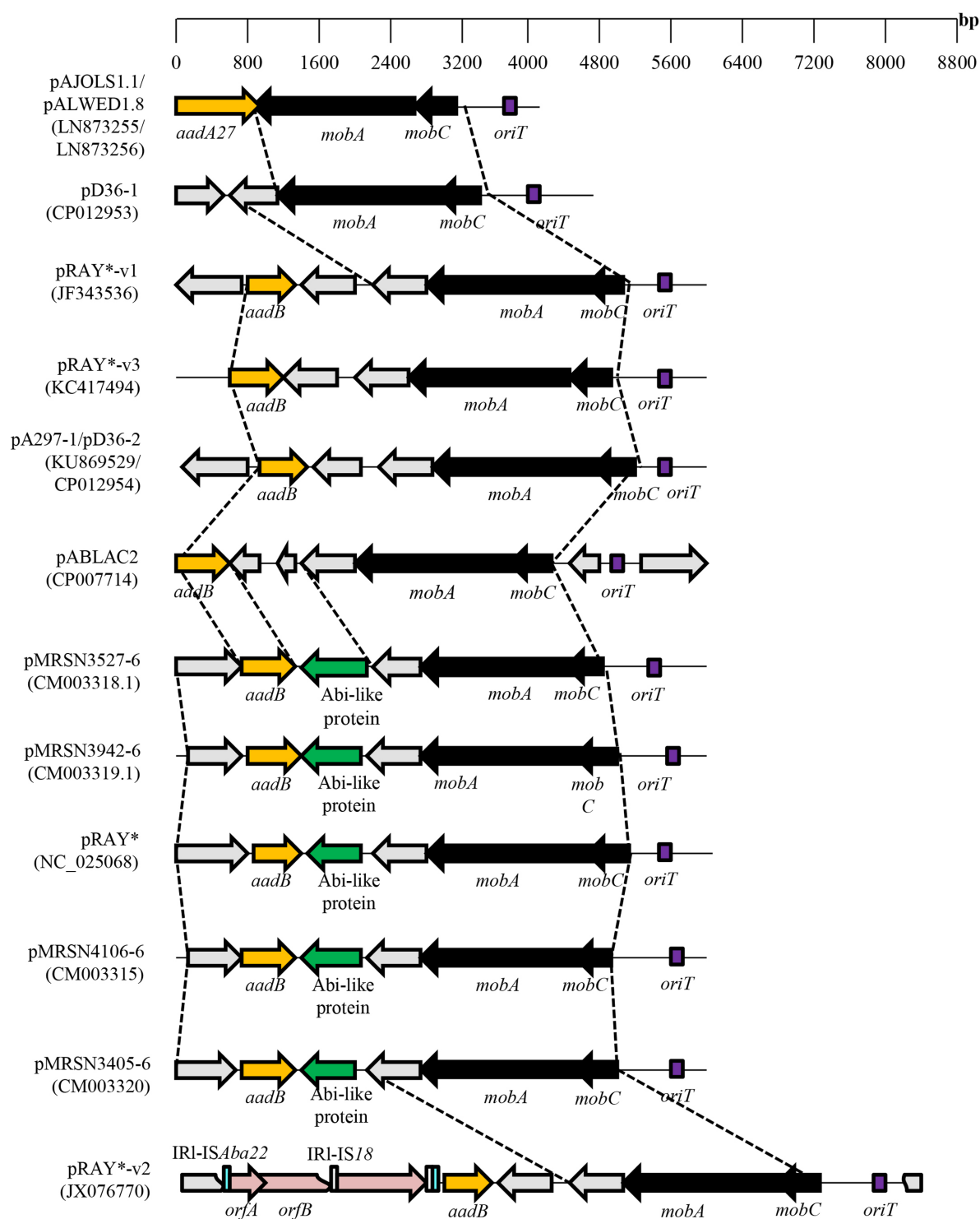
The 6,076 bp plasmid pRAY was first isolated from a South African clinical *Acinetobacter* strain designated SUN, which is of unknown clonal origin, through its carriage of the *aadB* gene which conferred resistance to the aminoglycosides gentamicin, kanamycin, and tobramycin (Segal and Elisha, 1999). The *aadB* gene is usually associated with class I integrons (Recchia and Hall, 1995), but in *Acinetobacter* sp. SUN and subsequently, in other *Acinetobacter* spp. isolated worldwide, *aadB* is found in pRAY and its closely related derivatives (Segal and Elisha, 1999; Adams et al., 2010; Nigro et al., 2011; Hamidian et al., 2012; Gifford et al., 2014; Ou et al., 2015; Kurakov et al., 2016). The *aadB* gene is likely acquired as its G+C content of 58% is higher than the G+C content of 37% for the rest of pRAY (Segal and Elisha, 1999) and the presence of an *attC* site immediately downstream of *aadB* is indicative of its gene cassette origin (Nigro et al., 2011).

A total of 10 ORFs, including *aadB*, was identified from the pRAY sequence, with two ORFs (designated ORF3 and ORF6) encoding proteins that were homologous to mobilization proteins (Segal and Elisha, 1999). A putative origin of transfer (*oriT*) was also identified upstream of ORF3 (Segal and Elisha, 1999), inferring the potential transmissibility of pRAY.

Derivatives of pRAY have been characterized from Australian *A. baumannii* clinical strains with a plasmid designated pRAY\* isolated from strain D36 and pRAY\*-v1 from strain C2 (Hamidian et al., 2012). The *mobA* gene from pRAY\* is larger than ORF3 of pRAY but is still categorized within the ColE1 superfamily of MobA proteins (MOB<sub>HEN</sub> family) with the putative *oriT* located upstream of *mobC* (Hamidian et al., 2012). Plasmid pRAY\*-v1 differed from pRAY\* by 66 single nucleotide differences, 65 of which were within the *mobC-mobA* region leading only to amino acid substitutions of MobC and MobA but without any frameshifts. *A. baumannii* E7 harbored pRAY\*-v2 which was 2.5 kb larger than pRAY and sequence analysis indicated complete identity with pRAY\* but with the insertion of two IS elements, an IS18-like element which is found within ISAb22 and located upstream of the *aadB* gene (Hamidian et al., 2012) (Figure 2). A single nucleotide variant of pRAY\*-v1, designated pRAY\*-v3, was isolated from a clinical strain of *A. nosocomialis* from Melbourne, Australia (Gifford et al., 2014).

Analysis of a 4,135 bp plasmid designated pALWED1.8 harbored in *A. lwoffii* isolated from the permafrost in Russia indicated conservation of the *oriT-mobC-mobA* region with pRAY and its derivatives (Figure 2) (Kurakov et al., 2016). The pALWED1.8 plasmid contained an *aadA27* gene downstream of *mobA* that conferred resistance to streptomycin/spectinomycin but without an *attC* site that was observed for *aadB* in pRAY and its variants (Nigro et al., 2011). Interestingly, the *oriT-mobC-mobA* backbone was identified from the genome sequence of various *Acinetobacter* species with various genes found in the





**FIGURE 2 |** Comparative map of the pRAY plasmid and its derivatives from various *Acinetobacter* spp. The *mobA* and *mobC* genes are indicated as black arrows whereas the AT-rich putative *oriT* sequence is indicated as a purple box. Antimicrobial resistance genes (either *aadA27* for streptomycin/spectinomycin resistance or *aadB* for aminoglycoside resistance) are depicted as a dark yellow arrow while the Abi-like protein (identified by pfam07751) gene is indicated as a green arrow. IS element-encoded transposases are depicted in pink, the inverted repeats (IRs) for ISAb22 are shown as blue rectangles whereas the IRs for IS18 are shown as white rectangles (for pRAY\*-v2). Hypothetical open reading frames are depicted as gray arrows. Accession numbers for the plasmids shown are in parentheses following their names.

accessory regions of these plasmids such as an alkyl sulfatase gene (involved in the degradation of surface-active substances such as sodium dodecylsulfate, or SDS) in the plasmid from *A. radioresistens* SK82 (Kurakov et al., 2016). Thus, members of this group of plasmids, including pRAY and pALWED1.8, might have originated from a common ancestor and independently acquired different genes into the accessory region of the plasmid. The mobilization of pALWED1.8 was demonstrated in conjugation experiments between *A. lwoffii* strain ED23-35 which contained pALWED1.8 and a large conjugative plasmid pKLH208 (Kholodii et al., 2004) and *A. baylyi* BD413rif as the recipient.

Intriguingly, until now, no potential replication initiation protein could be identified for pRAY and its derivatives based on sequence homology (Hamidian et al., 2012; Kurakov et al., 2016). Nevertheless, a potential origin of replication was identified for pRAY upstream of *aadB* where eight copies of an AT-rich repeat sequence, AAAAAATAT, were found (Segal and Elisha, 1999). The replication of these plasmids may mirror that of plasmids such as ColE1 which do not encode a *rep* gene since their replicon only consists of an *oriV* with the host RNA polymerase transcriptional machinery taking care of the melting of duplex DNA and synthesis of pre-primer RNA for replication initiation (Brantl, 2014; Thomas et al., 2017). Efforts to transform pRAY into *E. coli* were not successful, implying that pRAY and its derivatives might be specific for *Acinetobacter* (Segal and Elisha, 1999).

## CONCLUDING REMARKS

This mini-review has highlighted the small plasmids of *A. baumannii*, whether cryptic, resistance-related, or even mobilizable plasmids, and inferred the likely importance of these plasmids to their host. The potential of these small plasmids in transferring antibiotic resistance and possibly, even virulence genes, among *Acinetobacter* species should not be overlooked as their promiscuity could be comparable to that of larger plasmids and thus, would have a significant impact on the evolution of *A. baumannii*. The dearth of experimental studies with regards to these small *Acinetobacter* plasmids, given the importance of *A. baumannii* in the World Health Organization list of priority pathogens (World Health Organization, 2017), is indeed surprising and needs to be addressed. The PCR-based replicon typing (PBRT) scheme developed by Bertini et al. (2010) would probably need updating in view of an ever increasing amount of *A. baumannii* plasmid sequence data although their Rep-based classification scheme into different GR groupings appeared to be still valid with respect to the small plasmids. Nevertheless, plasmids of the pRAY-type would require another classification scheme due to the lack of a replicase protein. Other plasmid typing schemes such as plasmid multi-locus sequence typing (pMLST) and MOB classification based on plasmid mobility genes (Francia et al., 2004; Garcillán-Barcia et al., 2011) would be difficult to apply for these small *Acinetobacter* plasmids due to their lack of loci used in these typing schemes. There is clearly a need for us to accurately identify individual plasmids especially

in this era of big data and whole genome sequencing (Orlek et al., 2017; Thomas et al., 2017), tracking the movement of plasmids and understanding their dynamic evolution, and small plasmids should not escape from our consideration simply because of their size.

## AUTHOR CONTRIBUTIONS

SSL and CCY conceived, analyzed the data, wrote, edited, and approved this manuscript.

## FUNDING

This work was supported by provisions from grant FRGS/1/2016/SKK11/UNISZA/01/1 from the Malaysian Ministry of Higher Education to CCY.

## ACKNOWLEDGMENTS

Our thanks and appreciation to T. Venkova for her help in checking the iteron sequences and for her suggestions. Our gratitude also to A. Brahmavamsa for inspiring us with the title of this manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01547/full#supplementary-material>

**FIGURE S1** | Comparative map of the small *Acinetobacter* plasmids of the Rep-3 superfamily. The *repB* replicase gene is indicated as a dark blue filled arrow, the putative *repA* gene is depicted in light blue. Hypothetical open reading frames are shown as unfilled arrows whereas black arrows are for the *mobA/mobL* mobilization genes. Red crosses indicate the XerC/XerD recombination sites. Filled blue twin-triangles depict the iterons that make up the putative origin of replication, *oriV*. Accession numbers and further details of the plasmids are as in **Supplementary Table S1** with detailed iteron sequences and locations on the respective plasmids in Supplementary Data Sheet 1.

**FIGURE S2** | Phylogenetic tree of the small *Acinetobacter* plasmids of the Rep-1 superfamily based on the Rep protein sequences, analyzed and drawn using MEGA7 (Kumar et al., 2016). Protein sequences were aligned using MUSCLE (Edgar, 2004), evolutionary history was inferred using the Neighbor-Joining method and the optimal tree (with the sum of branch length = 3.01508040) is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The plasmids were grouped according to the GR classification scheme proposed by Bertini et al. (2010) and indicated here as GR14 and GR16 in different colored boxes. Accession numbers for the plasmids used in the analysis are as follows: p3AB5075 (NZ\_CP008709.1), pBL63.1 (NC\_006959.1), pM131-10 (NC\_025169.1), pAB49 (L77992.1), pMRSN7339-2.3 (NZ\_CM003313.1), p4ABAYE (NC\_010403.1), pMRSN58-2.7 (NZ\_CM003316.1), pA85-1 (NC\_025107.1), and pTS236 (NC\_016977.1). Note that pBL63.1 was isolated from *Bacillus licheniformis* and was included in the analysis based on the findings of Guglielmetti et al. (2005).

**TABLE S1** | Features of the Rep-3 superfamily group of *Acinetobacter* plasmids.

## REFERENCES

- Adams, M. D., Chan, E. R., Molyneux, N. D., and Bonomo, R. A. (2010). Genomewide analysis of divergence of antibiotic resistance determinants in closely related isolates of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54, 3569–3577. doi: 10.1128/AAC.00057-10
- Bertini, A., Poirel, L., Mugnier, P. D., Villa, L., Nordmann, P., and Carattoli, A. (2010). Characterization and PCR-based replicon typing of resistance plasmids in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54, 4168–4177. doi: 10.1128/AAC.00542-10
- Billington, S. J., Jost, B. H., and Songer, J. G. (2000). Thiol-activated cytolysins: structure, function and role in pathogenesis. *FEMS Microbiol. Lett.* 182, 197–205. doi: 10.1111/j.1574-6968.2000.tb08895.x
- Bonnin, R. A., Poirel, L., and Nordmann, P. (2012). AbaR-type transposon structures in *Acinetobacter baumannii*. *J. Antimicrob. Chemother.* 67, 234–236. doi: 10.1093/jac/dkr413
- Brantl, S. (2014). Plasmid replication control by antisense RNAs. *Microbiol. Spectr.* 2:PLAS-0001-2013. doi: 10.1128/microbiolspec.PLAS-0001-2013
- Bui, D., Ramiscal, J., Trigueros, S., Newmark, J. S., Do, A., Sherratt, D. J., et al. (2006). Differences in resolution of *mwr*-containing plasmid dimers mediated by the *Klebsiella pneumoniae* and *Escherichia coli* XerC recombinases: potential implications in dissemination of antibiotic resistance genes. *J. Bacteriol.* 188, 2812–2820. doi: 10.1128/JB.188.8.2812-2820.2006
- D'Andrea, M. M., Giani, T., D'Arezzo, S., Capone, A., Petrosillo, N., Visca, P., et al. (2009). Characterization of pABVA01, a plasmid encoding the OXA-24 carbapenemase from Italian isolates of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 53, 3528–3533. doi: 10.1128/AAC.00178-09
- Dorsey, C. W., Tomaras, A. P., and Actis, L. A. (2006). Sequence and organization of pMAC, an *Acinetobacter baumannii* plasmid harboring genes involved in organic peroxide resistance. *Plasmid* 56, 112–123. doi: 10.1016/j.plasmid.2006.01.004
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Fournier, P.-E., Vallenet, D., Barbe, V., Audic, S., Ogata, H., Poirel, L., et al. (2006). Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genet.* 2:e7. doi: 10.1371/journal.pgen.0020007
- Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., and De La Cruz, F. (2004). A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol. Rev.* 28, 79–100. doi: 10.1016/j.femsre.2003.09.001
- Gallagher, L. A., Ramage, E., Weiss, E. J., Radey, M., Hayden, H. S., Held, K. G., et al. (2015). Resources for genetic and genomic analysis of emerging pathogen *Acinetobacter baumannii*. *J. Bacteriol.* 197, 2027–2035. doi: 10.1128/JB.00131-15
- Garcillán-Barcia, M. P., Alvarado, A., and de la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol. Rev.* 35, 936–956. doi: 10.1111/j.1574-6976.2011.00291.x
- Gifford, B., Tucci, J., McIlroy, S. J., and Petrovski, S. (2014). Isolation and characterization of two plasmids in a clinical *Acinetobacter nosocomialis* strain. *BMC Res. Notes* 7:732. doi: 10.1186/1756-0500-7-732
- Göttig, S., Gruber, T. M., Higgins, P. G., Wachsmuth, M., Seifert, H., and Kempf, V. A. J. (2014). Detection of pan drug-resistant *Acinetobacter baumannii* in Germany. *J. Antimicrob. Chemother.* 69, 2578–2579. doi: 10.1093/jac/dku170
- Grosso, F., Quinteira, S., Poirel, L., Novais, A., and Peixe, L. (2012). Role of common *bla*<sub>OXA-24/OXA-40</sub>-carrying platforms, and plasmids in the spread of OXA-24/OXA-40 among *Acinetobacter* species clinical isolates. *Antimicrob. Agents Chemother.* 56, 3969–3972. doi: 10.1128/AAC.06255-11
- Guglielmetti, S., Mora, D., Manachini, P. L., and Parini, C. (2005). Genetic relationship among *Bacillus licheniformis* rolling-circle-replicating plasmids and complete nucleotide sequence of pBL63.1, an atypical replicon. *Plasmid* 54, 93–103. doi: 10.1016/j.plasmid.2005.01.002
- Hamidian, M., Ambrose, S. J., and Hall, R. M. (2016a). A large conjugative *Acinetobacter baumannii* plasmid carrying the *sul2* sulphonamide and *strAB* streptomycin resistance genes. *Plasmid* 8, 43–50. doi: 10.1016/j.plasmid.2016.09.001
- Hamidian, M., and Hall, R. M. (2014). *pACICU2* is a conjugative plasmid of *Acinetobacter* carrying the aminoglycoside resistance transposon *TnaphA6*. *J. Antimicrob. Chemother.* 69, 1146–1148. doi: 10.1093/jac/dkt488
- Hamidian, M., Holt, K. E., Pickard, D., Dougan, G., and Hall, R. M. (2014a). A GC1 *Acinetobacter baumannii* isolate carrying AbaR3 and the aminoglycoside resistance transposon *TnaphA6* in a conjugative plasmid. *J. Antimicrob. Chemother.* 69, 955–958. doi: 10.1093/jac/dkt454
- Hamidian, M., Holt, K. E., Pickard, D., and Hall, R. M. (2016b). A small *Acinetobacter* plasmid carrying the *tet39* tetracycline resistance determinant. *J. Antimicrob. Chemother.* 71, 269–271. doi: 10.1093/jac/dkv293
- Hamidian, M., Kenyon, J. J., Holt, K. E., Pickard, D., and Hall, R. M. (2014b). A conjugative plasmid carrying the carbapenem resistance gene *bla*<sub>OXA-23</sub> in AbaR4 in an extensively resistant GC1 *Acinetobacter baumannii* isolate. *J. Antimicrob. Chemother.* 69, 2625–2628. doi: 10.1093/jac/dku188
- Hamidian, M., Nigro, S. J., and Hall, R. M. (2012). Variants of the gentamicin and tobramycin resistance plasmid pRAY are widely distributed in *Acinetobacter*. *J. Antimicrob. Chemother.* 67, 2833–2836. doi: 10.1093/jac/dks318
- Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 301, 1496–1499. doi: 10.1126/science.1088157
- Jurenaite, M., Markuckas, A., and Suziedeliene, E. (2013). Identification and characterization of type II toxin-antitoxin systems in the opportunistic pathogen *Acinetobacter baumannii*. *J. Bacteriol.* 195, 3165–3172. doi: 10.1128/JB.00237-13
- Kholodii, G., Mindlin, S., Gorlenko, Z., Petrova, M., Hobman, J., and Nikiforov, V. (2004). Translocation of transposition-deficient (TndPKLH2-like) transposons in the natural environment: mechanistic insights from the study of adjacent DNA sequences. *Microbiology* 150(Pt 4), 979–992. doi: 10.1099/mic.0.26844-0
- Konieczny, I., Bury, K., Wawrzycka, A., and Wegrzyn, K. (2014). Iteron plasmids. *Microbiol. Spectr.* 2:PLAS-0026-2014. doi: 10.1128/microbiolspec.PLAS-0026-2014
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kurakov, A., Mindlin, S., Beletsky, A., Shcherbatova, N., Rakitin, A., Ermakova, A., et al. (2016). The ancient small mobilizable plasmid pALWED1.8 harboring a new variant of the non-cassette streptomycin/spectinomycin resistance gene *aadA27*. *Plasmid* 84, 36–43. doi: 10.1016/j.plasmid.2016.02.005
- Lean, S. S., Suhaili, Z., Ismail, S., Rahman, N. I. A., Othman, N., Abdullah, F. H., et al. (2014). Prevalence and genetic characterization of carbapenem- and polymyxin-resistant *Acinetobacter baumannii* isolated from a tertiary hospital in Terengganu, Malaysia. *ISRN Microbiol.* 2014:953417. doi: 10.1155/2014/953417
- Lean, S. S., Yeo, C. C., Suhaili, Z., and Thong, K.-L. (2015). Whole-genome analysis of an extensively drug-resistant clinical isolate of *Acinetobacter baumannii* AC12: insights into the mechanisms of resistance of an ST195 clone from Malaysia. *Int. J. Antimicrob. Agents* 45, 178–182. doi: 10.1016/j.ijantimicag.2014.10.015
- Lean, S. S., Yeo, C. C., Suhaili, Z., and Thong, K. L. (2016). Comparative genomics of two ST 195 carbapenem-resistant *Acinetobacter baumannii* with different susceptibility to polymyxin revealed underlying resistance mechanism. *Front. Microbiol.* 6:1445. doi: 10.3389/fmicb.2015.01445
- Li, M. S., Farrant, J. L., Langford, P. R., and Kroll, J. S. (2003). Identification and characterization of genomic loci unique to the Brazilian purpuric fever clonal group of *H. influenzae* biogroup aegyptius: functionality explored using meningococcal homology. *Mol. Microbiol.* 47, 1101–1111. doi: 10.1046/j.1365-2958.2003.03359.x
- Li, M.-S., Langford, P. R., and Kroll, J. S. (2017). Inactivation of NMB0419 encoding a Sell-like repeat (SLR) protein in *Neisseria meningitidis* is associated with differential expression of genes belonging to the Fur regulon and reduced intra-epithelial replication. *Infect. Immun.* 85:e00574-16. doi: 10.1128/IAI.00574-16
- Liu, C.-C., Tang, C. Y., Kuo, H.-Y., Lu, C.-W., Chang, K.-C., and Liou, M.-L. (2013). The origin of *Acinetobacter baumannii* TYTH-1: a comparative genomics study. *Int. J. Antimicrob. Agents* 41, 318–324. doi: 10.1016/j.ijantimicag.2012.12.010
- Magiorakos, A., Srinivasan, A., Carey, R., Carmeli, Y., Falagas, M., Giske, C., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281. doi: 10.1111/j.1469-0691.2011.03570.x

- Merino, M., Acosta, J., Poza, M., Sanz, F., Beceiro, A., Chaves, F., et al. (2010). OXA-24 carbapenemase gene flanked by XerC/XerD-like recombination sites in different plasmids from different *Acinetobacter* species isolated during a nosocomial outbreak. *Antimicrob. Agents Chemother.* 54, 2724–2727. doi: 10.1128/AAC.01674-09
- Meyer, R. (2009). Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* 62, 57–70. doi: 10.1016/j.plasmid.2009.05.001
- Midonet, C., and Barre, F.-X. (2014). Xer site-specific recombination: promoting vertical and horizontal transmission of genetic information. *Microbiol. Spectr.* 2:MDNA3-0056-2014. doi: 10.1128/microbiolspec.MDNA3-0056-2014
- Mittl, P. R. E., and Schneider-Brachert, W. (2007). Sell-like repeat proteins in signal transduction. *Cell. Signal.* 19, 20–31. doi: 10.1016/j.cellsig.2006.05.034
- Mosqueda, N., Gato, E., Roca, I., López, M., de Alegría, C. R., Fernández-Cuenca, F., et al. (2014). Characterization of plasmids carrying the *bla*<sub>OXA-24/40</sub> carbapenemase gene and the genes encoding the AbkA/AbkB proteins of a toxin/antitoxin system. *J. Antimicrob. Chemother.* 69, 2629–2633. doi: 10.1093/jac/dku179
- Nigro, S. J., Post, V., and Hall, R. M. (2011). Aminoglycoside resistance in multiply antibiotic-resistant *Acinetobacter baumannii* belonging to global clone 2 from Australian hospitals. *J. Antimicrob. Chemother.* 66, 1504–1509. doi: 10.1093/jac/dkr163
- Orlek, A., Stoesser, N., Anjum, M. F., Doumith, M., Ellington, M. J., Peto, T., et al. (2017). Plasmid classification in an era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology. *Front. Microbiol.* 8:182. doi: 10.3389/fmicb.2017.00182
- Ou, H.-Y., Kuang, S. N., He, X., Molgora, B. M., Ewing, P. J., Deng, Z., et al. (2015). Complete genome sequence of hypervirulent and outbreak-associated *Acinetobacter baumannii* strain LAC-4: epidemiology, resistance genetic determinants and potential virulence factors. *Sci. Rep.* 5:8643. doi: 10.1038/srep08643
- Peleg, A. Y., de Breij, A., Adams, M. D., Cerqueira, G. M., Mocali, S., Galardini, M., et al. (2012). The success of *Acinetobacter* species; genetic, metabolic and virulence attributes. *PLoS ONE* 7:e46984. doi: 10.1371/journal.pone.0046984
- Ramírez, M. S., Vilacoba, E., Stietz, M. S., Merckier, A. K., Jeric, P., Limansky, A. S., et al. (2013). Spreading of AbaR-type genomic islands in multidrug resistance *Acinetobacter baumannii* strains belonging to different clonal complexes. *Curr. Microbiol.* 67, 9–14. doi: 10.1007/s00284-013-0326-5
- Recchia, G. D., and Hall, R. M. (1995). Gene cassettes: a new class of mobile element. *Microbiology* 141, 3015–3027. doi: 10.1099/13500872-141-12-3015
- Segal, H., and Elisha, B. G. (1999). Characterization of the *Acinetobacter* plasmid, pRAY, and the identification of regulatory sequences upstream of an *aadB* gene cassette on this plasmid. *Plasmid* 42, 60–66.
- Thomas, C. M., Thomson, N. R., Cerdeño-Tárraga, A. M., Brown, C. J., Top, E. M., and Frost, L. S. (2017). Annotation of plasmid genes. *Plasmid* 91, 61–67. doi: 10.1016/j.plasmid.2017.03.006
- Val, M.-E., Bouvier, M., Campos, J., Sherratt, D., Cornet, F., Mazel, D., et al. (2005). The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Mol. Cell* 19, 559–566. doi: 10.1016/j.molcel.2005.07.002
- Wallace, L., Daugherty, S. C., Nagaraj, S., Johnson, J. K., Harris, A. D., and Rasko, D. A. (2016). Use of comparative genomics to characterize the diversity of *Acinetobacter baumannii* surveillance isolates in a health care institution. *Antimicrob. Agents Chemother.* 60, 5933–5941. doi: 10.1128/AAC.00477-16
- World Health Organization (2017). *WHO Publishes List of Bacteria for Which New Antibiotics are Urgently Needed*. Geneva: World Health Organization. Available at: <http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/>
- Zimble, D. L., Arivett, B. A., Beckett, A. C., Menke, S. M., and Actis, L. A. (2013). Functional features of TonB energy transduction systems of *Acinetobacter baumannii*. *Infect. Immun.* 81, 3382–3394. doi: 10.1128/IAI.00540-13

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Lean and Yeo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Fic Proteins of *Campylobacter fetus* subsp. *venerealis* Form a Network of Functional Toxin–Antitoxin Systems

Hanna Sprenger<sup>1,2,3†</sup>, Sabine Kienesberger<sup>1,2,4†</sup>, Brigitte Pertschy<sup>1</sup>, Lisa Pörtl<sup>1</sup>, Bettina Konrad<sup>1</sup>, Priya Bhutada<sup>1</sup>, Dina Vorkapic<sup>1</sup>, Denise Atzmüller<sup>1</sup>, Florian Feist<sup>5</sup>, Christoph Högenauer<sup>3</sup>, Gregor Gorkiewicz<sup>2,4</sup> and Ellen L. Zechner<sup>1,4\*</sup>

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Damian Lobato-Marquez,  
Imperial College London,  
United Kingdom  
Ramon Diaz Orejas,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### \*Correspondence:

Ellen L. Zechner  
ellen.zechner@uni-graz.at

† These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 04 July 2017

Accepted: 25 September 2017

Published: 17 October 2017

### Citation:

Sprenger H, Kienesberger S,  
Pertschy B, Pörtl L, Konrad B,  
Bhutada P, Vorkapic D, Atzmüller D,  
Feist F, Högenauer C, Gorkiewicz G  
and Zechner EL (2017) Fic Proteins  
of *Campylobacter fetus* subsp.  
*venerealis* Form a Network  
of Functional Toxin–Antitoxin  
Systems. *Front. Microbiol.* 8:1965.  
doi: 10.3389/fmicb.2017.01965

<sup>1</sup> Institute of Molecular Biosciences, University of Graz, Graz, Austria, <sup>2</sup> Institute of Pathology, Medical University of Graz, Graz, Austria, <sup>3</sup> Division of Gastroenterology and Hepatology, Medical University of Graz, Graz, Austria, <sup>4</sup> BioTechMed-Graz, Graz, Austria, <sup>5</sup> Vehicle Safety Institute, Graz University of Technology, Graz, Austria

Enzymes containing the FIC (filamentation induced by cyclic AMP) domain catalyze post-translational modifications of target proteins. In bacteria the activity of some Fic proteins resembles classical toxin–antitoxin (TA) systems. An excess of toxin over neutralizing antitoxin can enable bacteria to survive some stress conditions by slowing metabolic processes and promoting dormancy. The cell can return to normal growth when sufficient antitoxin is present to block toxin activity. *Fic* genes of the human and animal pathogen *Campylobacter fetus* are significantly associated with just one subspecies, which is specifically adapted to the urogenital tract. Here, we demonstrate that the *fic* genes of virulent isolate *C. fetus* subsp. *venerealis* 84-112 form multiple TA systems. Expression of the toxins in *Escherichia coli* caused filamentation and growth inhibition phenotypes reversible by concomitant antitoxin expression. Key active site residues involved in adenylation by Fic proteins are conserved in Fic1, Fic3 and Fic4, but degenerated in Fic2. We show that both Fic3 and the non-canonical Fic2 disrupt assembly and function of *E. coli* ribosomes when expressed independently of a trans-acting antitoxin. Toxicity of the Fic proteins is controlled by different mechanisms. The first involves intramolecular regulation by an inhibitory helix typical for Fic proteins. The second is an unusual neutralization by heterologous Fic–Fic protein interactions. Moreover, a small interacting antitoxin called Fic inhibitory protein 3, which appears unrelated to known Fic antitoxins, has the novel capacity to bind and neutralize Fic toxins encoded in *cis* and at distant sites. These findings reveal a remarkable system of functional crosstalk occurring between Fic proteins expressed from chromosomal and extrachromosomal modules. Conservation of *fic* genes in other bacteria that either inhabit or establish pathology in the urogenital tract of humans and animals underscores the significance of these factors for niche-specific adaptation and virulence.

**Keywords:** post-translational modification, adenylation, toxin–antitoxin module, bacterial effector protein, bacterial evolution, niche adaptation, urogenital tract, virulence

## INTRODUCTION

The genus *Campylobacter* comprises ecologically diverse species that colonize humans and animals. *Campylobacter jejuni* is known as the leading cause of human bacterial diarrhea worldwide. Other *Campylobacter* species, including *Campylobacter fetus*, are increasingly recognized as important human and animal pathogens (Lastovica and Allos, 2008; Man, 2011; Bullman et al., 2013). *C. fetus* is intriguing because although the two subspecies associated with mammals, *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis*, are highly related at the genome level, they exhibit quite different niche adaptations. *C. fetus* subsp. *fetus* has a broad host range (Skirrow and Benjamin, 1980; Harvey and Greenwood, 1985; Logue et al., 2003). In humans it causes gastrointestinal disease and belongs to the *Campylobacter* spp. most frequently associated with bacteremia (Lastovica and Allos, 2008; Man, 2011). By contrast, *C. fetus* subsp. *venerealis* is a host-restricted veterinary pathogen adapted to the urogenital tract of cattle (Blaser et al., 2008). Current understanding of the pathogenesis of emerging *Campylobacter* spp. is quite limited.

Recent comparative genomics of *C. fetus* subspecies revealed genetic determinants potentially contributing to this species' niche preferences and pathogenicity (Kienesberger et al., 2014; Graaf-van Bloois et al., 2016). Strikingly, *C. fetus* genomes encode multiple bacterial type IV secretion systems (T4SS), which generally contribute to pathogenicity by transferring specific protein and DNA substrates to recipient cells (Christie et al., 2014). One *C. fetus* T4SS has been evaluated experimentally and linked to virulence (Gorkiewicz et al., 2010). The conserved T4SS-encoding regions of *C. fetus* genomes fit into three phylogenetically different groups: one located exclusively on the chromosome, one observed exclusively on plasmids and a third located on both (Graaf-van Bloois et al., 2016). These authors further showed that both genes encoding T4SS components and genes encoding FIC domain proteins are significantly associated with the *C. fetus* subsp. *venerealis*. In the current study we focus on the function of the *fic* genes.

The Fido domain superfamily is composed of members of the FIC (filamentation induced by cAMP) and the Doc (death on curing) protein families and is common in all domains of life (Kinch et al., 2009). Proteins of the combined family contain a conserved motif [HPFx(D/E)GN(G/K)R]. Work in recent years has revealed that enzymes of the family catalyze post-translational modifications of proteins by addition of AMP, other nucleoside monophosphates, phosphocholine, or phosphate to a functionally critical amino acid (as reviewed in Cruz and Woychik, 2014; Garcia-Pino et al., 2014; Roy and Cherfils, 2015; Harms et al., 2016b). Since activities of the target proteins are typically altered as a result, Fido proteins are recognized as important regulators of metabolic functions.

Phylogenetic analysis of the superfamily places the paradigm Doc toxin of bacteriophage P1 in subfamily I (Garcia-Pino et al., 2014). Doc toxin is structurally similar to Fic proteins (Garcia-Pino et al., 2008); however, variation in the catalytic motif (K in place of the second G) confers kinase activity in place of NMP transfer activity (Castro-Roa et al., 2013; Cruz et al., 2014). Phosphorylation of translation elongation factor Tu by Doc leads

to rapid translation arrest in *Escherichia coli* (Garcia-Pino et al., 2008; Liu et al., 2008; Castro-Roa et al., 2013).

Interest in the FIC protein subfamily has been fueled by the observation that bacterial pathogens secrete Fic enzymes to modify host proteins (Pan et al., 2008; Worby et al., 2009; Yarbrough et al., 2009; Mukherjee et al., 2011). Cell to cell transfer can be direct via type III or type IV secretion (Roy and Cherfils, 2015). In the host, Fic effector proteins contribute to bacterial pathogenicity by modifying proteins important to signaling (Roy and Mukherjee, 2009; Woolery et al., 2010). Fic effectors contain the canonical catalytic motif [HxFx(D/E)GNRxxR] and initial studies showed that a typical reaction inactivates host GTPases by nucleotidyl transfer to a hydroxyl group of the protein side chain (Garcia-Pino et al., 2014; Roy and Cherfils, 2015). Several secreted FIC proteins transfer AMP in a reaction called adenylation (Worby et al., 2009; Yarbrough et al., 2009; Zekarias et al., 2010; Palanivelu et al., 2011), but variation within the canonical core motif can alter enzyme activity (Mukherjee et al., 2011; Engel et al., 2012).

The targets of FIC enzyme modification are not restricted to proteins expressed by the host. However, their functions and regulation in producing bacteria are still poorly understood. *E. coli* has been used as a surrogate producer to gain insights into the activities of FIC proteins in bacteria. One function that has emerged from these studies is that Fic proteins act as toxin-antitoxin (TA) modules (Harms et al., 2016b). Bacterial TA systems play a major role in cellular adaptation to stress and persistence (Hayes and Van Melderren, 2011; Goeders and Van Melderren, 2014; Harms et al., 2016a). Activation of the toxin can cause slow cell growth or arrest the cell cycle allowing bacteria to enter a dormant state. Mechanistic understanding of TA activity has been developed with prototypic modules such as *phd-doc* of bacteriophage P1 (Lehnher et al., 1993; Castro-Roa et al., 2013). Generally the toxin component is directed against the producing cell and interferes with bacterial physiology. Cellular processes inhibited by type II TA toxins include protein synthesis, cell wall synthesis, assembly of cytoskeletal structures and DNA topoisomerase action (Hayes and Van Melderren, 2011; Yamaguchi and Inouye, 2011; Goeders and Van Melderren, 2014; Harms et al., 2016b). The antitoxin component reversibly inactivates the toxin and/or regulates its expression. Unlike the toxin, the antitoxin is biochemically unstable so that, unless the antitoxin is continuously expressed, the free toxin forces the bacterial cell into a reversible dormant state or even kills the cell (Leplae et al., 2011; Goeders and Van Melderren, 2014). The TA system of bacteriophage P1 helps to maintain the lysogen through post-segregational killing of cells that are cured of the prophage (Lehnher et al., 1993). Homologous *phd-doc* modules are also present on bacterial chromosomes and evidence thus far suggests a role for these systems in the formation of persister cells under stress conditions (Maisonneuve and Gerdes, 2014). Moreover, evidence is emerging that TA modules help bacteria overcome stress imposed by host colonization, early stages of infection and survival within host cells (Norton and Mulvey, 2012; Ren et al., 2012, 2014; De la Cruz et al., 2013; Lobato-Marquez et al., 2015). They can stabilize mobile genetic elements encoding virulence factors and contribute directly to

virulence (see Lobato-Marquez et al., 2016 for a comprehensive review).

*Campylobacter* genomes generally lack homologs of prototypical TA systems (Shao et al., 2011). To date, only two TA systems (both located on a plasmid in *C. jejuni*) have been described (Shen et al., 2016). Given the general importance of TA systems in bacteria we asked whether the multiple *fic* genes in *C. fetus* fulfill this important role. Here, we show that the Fic proteins of *C. fetus* subsp. *venerealis* 84-112 indeed form TA systems with the capacity to disrupt the bacterial translational machinery. We further show that *fic* modules located on the chromosome and extrachromosomal DNA functionally interact. *Fic* homologs are genetically conserved in *C. fetus* subsp. *venerealis* isolates and in other human and animal urogenital pathogens, underscoring the significance of these factors for niche-specific adaptation.

## MATERIALS AND METHODS

### Bacteria

Strains used in this study are listed in Supplementary Table S1. *E. coli* and *Campylobacter* strains were grown as previously described (Kienesberger et al., 2007). Antibiotics were added to final concentrations of 100  $\mu\text{g ml}^{-1}$  ampicillin, 75  $\mu\text{g ml}^{-1}$  nalidixic acid, and either 12.5 or 25  $\mu\text{g ml}^{-1}$  chloramphenicol, 40 or 25  $\mu\text{g ml}^{-1}$  kanamycin for *E. coli* or *Campylobacter* cultivation, respectively.

### Construction of Plasmids

Plasmids and oligonucleotides used in this study are listed in Supplementary Tables S1, S2. For expression in *E. coli*, genes of interest were amplified with PCR and the fragments were ligated to pBAD24 vector derivatives with distinct antibiotic resistance genes (see Supplementary Tables S1, S2).

### Structure Predictions

For 3D structure prediction amino acid sequences (CDF65254.1, CDF65253.1, CDF65920.1, and CDF65967.1) were analyzed using the Phyre2 web portal (Rollins and Colwell, 1986). The output files were rendered with PyMOL (Schrodinger, 2010). Templates for Fic protein fold recognition via Phyre2 are listed in Supplementary Table S3.

### *E. coli* Growth/Rescue Assays

*Escherichia coli* DH5 $\alpha$  harboring pBAD plasmids with *fic* genes were grown with shaking overnight at 37°C in LB-broth supplemented with appropriate antibiotics and 0.2% glucose to repress expression, or 0.05% arabinose to induce expression via the  $P_{\text{BAD}}$  promoter. Bacterial growth was either monitored by survival plating of bacteria grown in 100 ml LB-broth or in 24-well plates with a culture volume of 1 ml per well and starting optical density measured at 600 nm ( $\text{OD}_{600}$ ) of 0.05. Plates were incubated at 37°C under shaking at 180 rpm.  $\text{OD}_{600}$  was measured hourly in triplicate. Determination of colony forming units (CFUs) normally corresponded with  $\text{OD}_{600}$  measurements

except at late time points where the CFU count of filamentous cells remained low.

### Microscopy

Cultures of *E. coli* DH5 $\alpha$  harboring pBAD derivatives with a starting  $\text{OD}_{600}$  of 0.05 were grown in LB-broth with 0.05% arabinose for 2 h. Cells were harvested, suspended in 1x phosphate buffered saline (PBS), pH 7.4 and incubated with Nile red for up to 60 min at room temperature in the dark. For immediate microscopy, 1  $\mu\text{l}$  of the pellet was applied to an agar slide (1% agar solution poured on microscopy slide) to immobilize the cells. For later imaging, cells were fixed with 0.4% formaldehyde, before collected by centrifugation, resuspended in 1x PBS, pH 7.4 and stained as described before. Confocal microscopy was performed on a LEICA AOBSP2 MP microscope (380 nm extinction, 510 nm emission).

### Co-immunoprecipitation and Western Analysis

Pairs of FLAG-tagged and hemagglutinin (HA)-tagged proteins were co-expressed in *E. coli* C41(DE3) from respective plasmids (Supplementary Table S1). 100 ml LB broth supplemented with 0.2% glucose was inoculated with overnight cultures to  $\text{OD}_{600}$  of 0.1. When cultures reached  $\text{OD}_{600} = 0.5\text{--}0.8$  protein expression was induced with 0.05% arabinose for 2 h. Thirty OD of cells were pelleted and washed with 50 ml buffer A (50 mM Tris-HCl pH 6.8, 100 mM NaCl). Cell lysis was performed as previously described (Gruber et al., 2016) except that the formaldehyde crosslinking step was omitted for this study. All further steps were as in Gruber et al. (2016). For protein detection,  $\text{OD}_{600}$  0.015–0.05 equivalents of lysate and pull-down fractions were mixed with sample buffer containing DTT (0.09%) and SDS (0.1%), heated at 95°C for 10 min and resolved on SDS-PAGE (12.5%, Hoefer) or NUPAGE (12%, MES buffer, Invitrogen) gels. Proteins were transferred for 1.5 h onto PVDF membranes. Blocking was done overnight at 4°C in TST: (0.5 M Tris-HCl pH 7.5, 1.5 M NaCl, 1% Tween-20) supplemented with 3% milk powder. FLAG-tagged proteins were detected with HRP-conjugated  $\alpha$ -FLAG antibody (A8592, Sigma) and HA-tagged proteins with HRP-conjugated  $\alpha$ -HA antibody (12013819001, Roche). After washing (3 min  $\times$  10 min) with 1x TST blots were developed with ECL (Bio-Rad) according to the manufacturer's instructions (1:5 dilution in ddH<sub>2</sub>O of substrate for FLAG detection).

### Ribosome Profiles

*Escherichia coli* DH5 $\alpha$  with plasmids were grown in 100 ml LB-broth to an  $\text{OD}_{600}$  of 0.4 to 0.6, then shifted to medium containing 0.05% arabinose for 1 h to induce expression. 30 s prior to cell harvest, chloramphenicol was added to a final concentration of 200 or 300  $\mu\text{g ml}^{-1}$ . Cells were harvested by centrifugation for 10 min at 14,300  $\times g$  and 4°C. The cell pellet was resuspended in 500  $\mu\text{l}$  cell lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 30 mM NH<sub>4</sub>Cl, 100 or 150  $\mu\text{g ml}^{-1}$  chloramphenicol) and either mixed with an equal volume of glass beads (300  $\mu\text{m}$  in diameter) and vortexed for 5 min at 4°C or



immediately frozen in liquid nitrogen (Bronowski et al., 2014). Suspensions with glass beads were centrifuged for 10 min at  $6,400 \times g$  at  $4^{\circ}\text{C}$ . The supernatant was collected, centrifuged for 3 min at  $17,649 \times g$  at  $4^{\circ}\text{C}$  and immediately applied to sucrose density centrifugation or stored at  $-70^{\circ}\text{C}$ . Frozen suspensions were thawed in an ice bath, frozen again in liquid nitrogen and stored at  $-70^{\circ}\text{C}$ .

The protocol for the sucrose density centrifugation was adapted from Jiang et al. (2007). An  $A_{260}$  of 8 of the cleared cell lysate was loaded onto a gradient of 5–45% sucrose in buffer (10 mM Tris-HCl pH 7.5, 10 mM  $\text{MgCl}_2$ , 100 mM  $\text{NH}_4\text{Cl}$ ). Ultracentrifugation was carried out for 4 h at  $4^{\circ}\text{C}$  and  $253,483 \times g$  in a Beckman SW-41Ti rotor. Fractions of the gradient were collected using an UA-6 system (Teledyne ISCO) with continuous monitoring at  $A_{254}$ .

## Numerical and Statistical Analysis of Ribosome Profiles

Ribosome profiles were scanned, traced in CorelDraw to increase contrast and xy-coordinates were extracted using DataThief III (Pascoe et al., 2015). To increase the reliability of calculations, in addition to peak values we calculated the Xgrad-values using a code (XSpan) written in VisualBasic. The code is available to interested readers upon request. The program 'XSpan' places the largest rectangular surface with a predefined width (= X) under individual peaks of a given curve and then calculates the Xgrad-values, which are height (H) and area (A) of the surfaces. XSpan can also extrapolate clipped curves (e.g., when a maximum value exceeds the measurement range) by fitting a cubic function such that it tangents the two flanks of a clipped peak. This option was utilized in this study to obtain the 70S heights (H). The Xgrad-value H is the highest amplitude (absorbance) measured for the surface of predefined width describing fractions of the analyzed sedimentation gradient. X, the width of the rectangular surfaces, was selected such that it covers approximately 1.2% of the gradient fractions analyzed. Statistical analysis was performed using the H values only. For each profile, the H values of the 30S and 50S subunit peaks were normalized to the H values of the first polysome peak and statistical significance was calculated using the paired Student's *t*-test. Statistical significance was assumed with *p*-values below 0.05.

## PCR Screening, DNA Sequencing, and Sequence Alignment of *fic* Genes

Prevalence of TA genes was surveyed among *C. fetus* isolates via PCR using chromosomal DNA as template. We applied primer pairs 1/2 for *fic1*, 26/27 for *fic2*, 5/6 for *fic3*, 7/8 for *fic4*, and 9/10 for *fti3* (Supplementary Table S2). Sequencing of *fic2* amplicons from *C. fetus* subsp. *venerealis* strains V9, V20, V32, V60, V62, and V69 was performed with primers 28/29.

## Phylogenetic Analysis of *Campylobacter* spp. FIC Proteins

The conserved Fido motif sequence HPFXXGNXR and full length Fic1-4 of *C. fetus* subsp. *venerealis* 84-112 were used in BlastP analysis to identify Fido proteins in whole genomes of

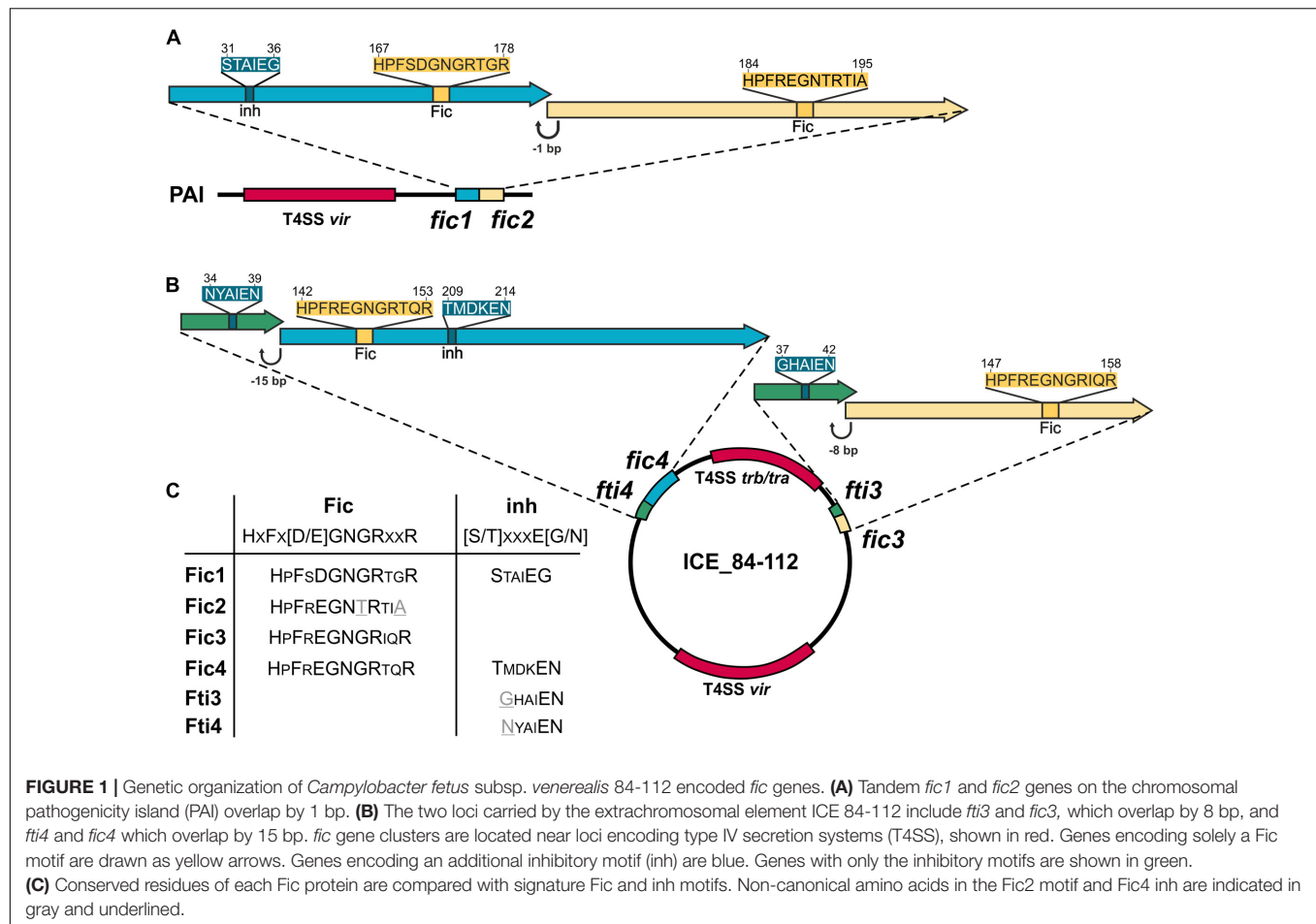
*Campylobacter* species (if possible, finished whole genomes, if not available, genomes with low scaffold numbers were used). BlastP analysis was also performed with full-length Fics and the Fic2 specific motif sequence [HPFREGNTRTIA] under exclusion of epsilon-proteobacteria to screen for hits outside this class. Selected *Campylobacter* proteins, selected Fic reference proteins, as well as other bacteria from the urogenital tract identified by the BlastP were then used to generate the phylogeny tree. Retrieved proteins were aligned with MEGA6.06 using the BLOSUM matrix. The Neighbor joining tree was constructed with MEGA6.06 (Tamura et al., 2013). The tree was rooted to the translated ORF of housekeeping gene *glnA* of *C. fetus* subsp. *venerealis* 84-112. Protein accession numbers are listed in Supplementary Table S4.

## RESULTS

### FIC Domain Proteins of *C. fetus* subsp. *venerealis* 84-112

*fic1* and *fic2* genes (Figure 1A) are chromosomally encoded and form part of a pathogenicity island (PAI) that harbors additionally a functional T4SS (Gorkiewicz et al., 2010). *C. fetus* subsp. *venerealis* 84-112 also carries extra-chromosomal DNA with features of an integrative conjugative element (ICE\_84-112) (Kienesberger et al., 2014). Two additional *fic* gene homologs, *fic3* and *fic4*, were identified on the ICE (Figure 1B). Residues of the Fido superfamily core motif that enable FIC-containing enzymes to act as AMP transferases have been defined as HxFx(D/E)GNGRxxR (Kinch et al., 2009; Worby et al., 2009; Xiao et al., 2010; Engel et al., 2012). Fic1, Fic3, and Fic4 contain the complete signature of invariant residues [protein accession numbers CDF65254.1 (Fic1); CDF65920.1 (Fic3); CDF65967.1 (Fic4)]. In contrast, in Fic2 (CDF65253.1), the second conserved glycine at position 191 is replaced with threonine and the final arginine of the signature motif (R195A) is absent, suggesting that Fic2 does not have adenylation activity. Fic1 and Fic4 also contain a conserved inhibitory motif (S/T)xxxE(G/N), which was shown to suppress adenylation in well-studied systems (Harms et al., 2016b) (Figure 1C). FIC proteins containing this inhibitory helix (inh) are classified depending on whether the inh is part of the FIC fold as an N-terminal helix (class II) or a C-terminal helix (class III) (Engel et al., 2012). Fic1 thus belongs to class II, and Fic4 to class III. Class I FIC proteins do not contain the inhibitory motif themselves, but have an interaction partner that provides the inh *in trans*. Fic2 and Fic3 lack a motif with this overall consensus, thus they belong to class I. Fic1 may act as antitoxin for the degenerated toxin Fic2. Alternatively, mutations of the core motif in Fic2 may have altered enzyme activity and thus have bypassed the need for an inh motif. We also note that the 78 amino acid ORF (protein accession number CDF65919.1) upstream and partially overlapping *fic3* includes residues GHAIEN, which might provide the invariable glutamate (Engel et al., 2012; Goepfert et al., 2013) of a poorly conserved inhibitory motif (Figure 1B). Presence of a 48% identical homolog, *fti4* (protein accession number CDF65966.1), upstream and partially overlapping *fic4* strengthens the hypothesis that





each ORF encodes a small interacting protein to control the cognate FIC enzyme.

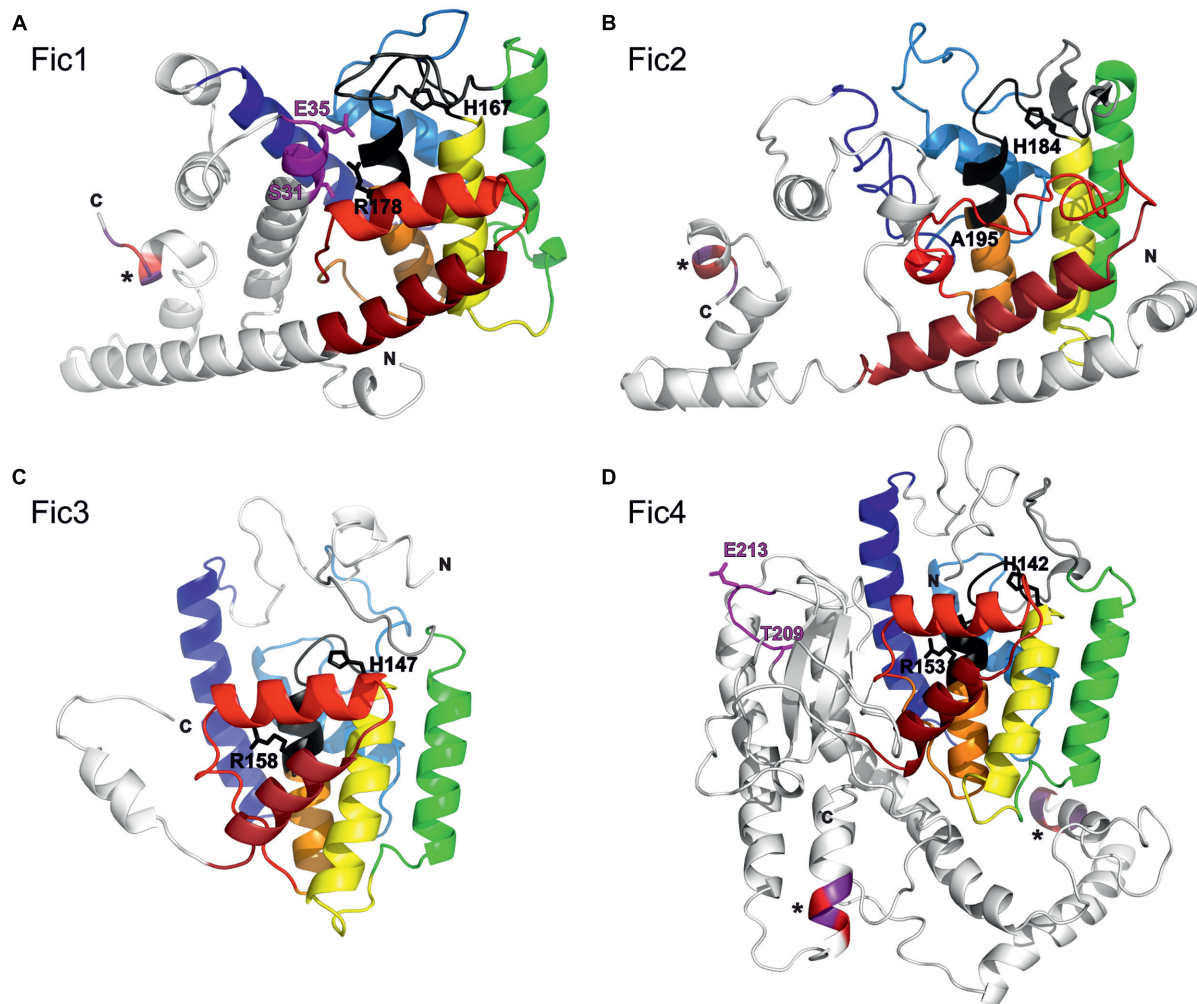
## Protein Structure Prediction

Alignment of the predicted proteins shows strong conservation of the Fic core motif but low general similarity (not shown). A structure prediction was performed with the Phyre2 server using templates listed in Supplementary Table S4. All *C. fetus* subsp. *venerealis* homologs are predicted to share a similar FIC domain fold (**Figures 2A–D**). The set of common  $\alpha$ -helices are colored from blue (N-terminal) to red (C-terminal), according to the core FIC domain secondary structure topology (Kinch et al., 2009). The predicted active site loops with the conserved core motif including the catalytic histidine are highlighted in black. Fic proteins typically carry a  $\beta$  hairpin close to the active site (gray). This structure, also called “the flap,” constitutes the major target-protein docking site (Kinch et al., 2009; Xiao et al., 2010; Palanivelu et al., 2011; Garcia-Pino et al., 2014). The inhibitory motifs of Fic1 and Fic4, expected to prevent the adenylation reaction by active site obstruction, are shown in pink. The remaining protein structure outside of each FIC core domain is shown in white. One additional shared feature we noted is the conserved KEKE motif (asterisks) at the C-termini of Fic1 and Fic2 that is reiterated in Fic4, once internally, and again at the

C-terminus. Bacterial effector proteins secreted via a given T4SS typically display a short C-terminal stretch of conserved residues that mediates their specific recognition by the transfer machinery (Zechner et al., 2012; Christie et al., 2014). The conserved KEKE motif may represent such a dedicated translocation signal, but this has not been validated experimentally.

## Fic1 and Fic2 Form a Functional TA Module

To gain insights to the function of the *C. fetus* proteins we expressed these in a heterologous bacterial host and asked whether the *fic1-fic2* module acts as a TA system. In that case, the inhibitory domain of Fic1 would be required to act both intra- and intermolecularly to regulate the enzymatic activity of Fic1 and Fic2. The *fic* genes of *C. fetus* were placed under transcriptional control of the  $P_{BAD}$  promoter and their effects on growth of *E. coli* were investigated. Shifting *E. coli* cells from LB broth with glucose to medium containing arabinose induced synthesis of the Fic proteins and culture density was monitored over time. Induction of *fic2* expression delayed growth of *E. coli* severely compared to the vector control, demonstrating that Fic2 is toxic despite its degenerate core motif (**Figure 3A**). Exchange of the catalytic histidine in variant Fic2\_H184A eliminated

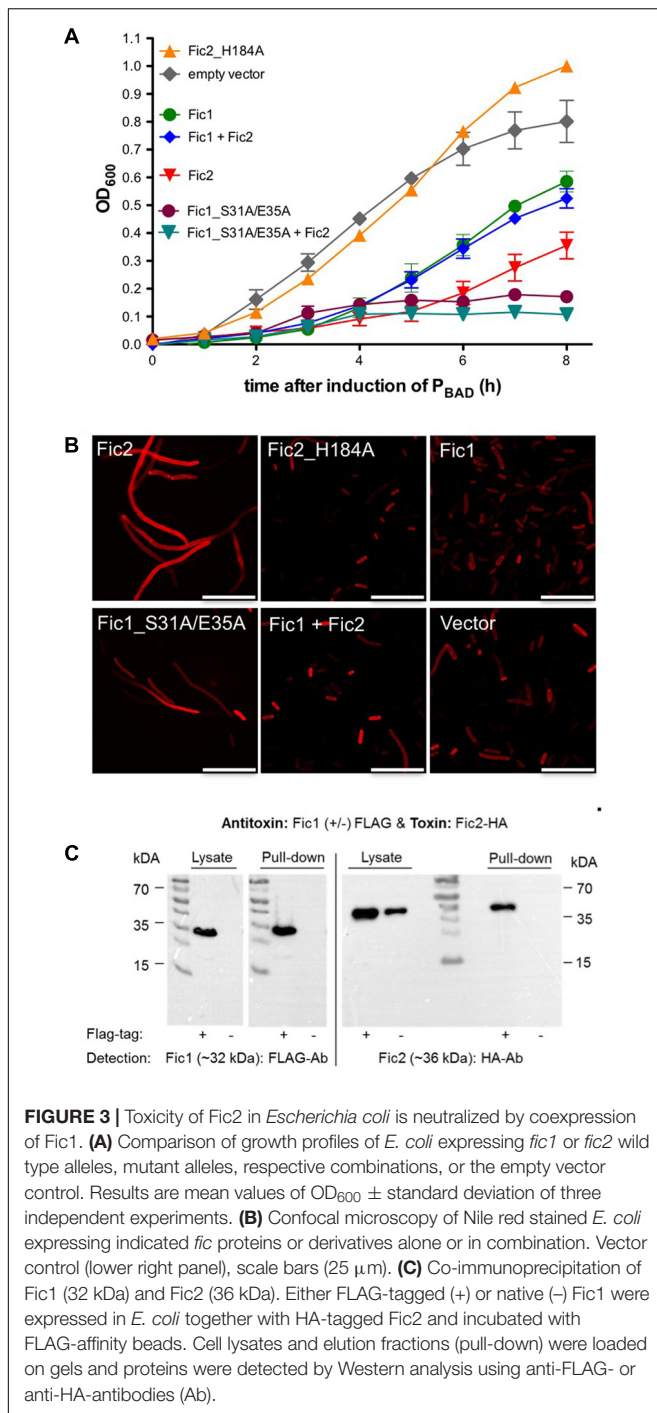


**FIGURE 2 |** Structural models of *C. fetus* subsp. *venerealis* Fic proteins. The structure of each Fic protein (A–D) was predicted with Phyre2. The rainbow colors from blue (N-termini; N) to red (C-termini; C) illustrate the core FIC domain secondary structure topology, as defined in Kinch et al. (2009). Black designates each FIC core motif. Pink specifies the inhibitory motifs (Fic1 and Fic4). Key residues within these motifs are numbered. Protein extensions outside the Fic fold are shown in white. Asterisk designates putative C-terminal translocation signals containing conserved KEKE residues.

toxicity and allowed the host to grow comparably to the vector control strain. Loss of phenotype could occur either because the histidine is indeed important to the activity of the enzyme, as predicted, or because the mutant variant is unstable. To exclude the latter possibility we purified the mutated protein and verified its stability during overexpression in *E. coli* and in isolated form (not shown). *E. coli* expressing *fic1* displayed logarithmic growth, but culture densities obtained after 8 h were lower than cells carrying the empty vector. To examine the role of the conserved inh of Fic1, key residues Ser31 and Glu35 were exchanged for alanine. The substitution apparently disrupted the protective function of this motif, since expression of Fic1\_S31A/E35A was incompatible with cell growth. Cells were rescued from Fic2-induced growth arrest by co-expression of wild type Fic1, suggesting that Fic1 can act as an antitoxin for Fic2. The importance of the inh module in toxin neutralization

was again shown when co-expression of Fic1\_S31A/E35A and Fic2 arrested growth fully. The data imply that Fic1 catalyzes an activity detrimental to bacterial growth, but which is normally blocked intramolecularly by the protein's inh helix. Moreover, the bacterial cytotoxicity of Fic2 depends on the enzyme core motif and is neutralized by antitoxin Fic1.

To characterize the proposed toxin–antitoxin activities, we next compared the impact of Fic protein production on cellular morphology. Fic2 alone caused an extreme filamentous phenotype (Figure 3B) via a mechanism requiring the catalytic histidine since; by comparison, cells expressing Fic2H184A were similar to wild type. Cells expressing Fic1 appeared normal but formed filaments when the Fic1 inh motif was mutated. Coexpression of wild type Fic1 reversed the filamentous phenotype caused by Fic2 consistent with the neutralization observed during growth (Figure 3A).



Antitoxins similar to the Phd-Doc paradigm frequently inactivate the toxin by forming a stable complex. We asked whether inactivation of Fic2 toxicity by Fic1 might involve binding of the two proteins. Codons for a FLAG epitope were added to *fic1* and the hemagglutinin (HA) tag was added to *fic2*. Lysates of *E. coli* cells expressing both fusion proteins were incubated with FLAG-affinity beads. After elution of bound proteins, lysates and eluates were analyzed by western immunoblotting (Figure 3C). Anti-FLAG antibodies confirmed

the presence of FLAG-tagged Fic1 in cell lysates and the absence of signal in control samples expressing native Fic1. Antibody to HA detected Fic2-HA fusion protein in the same cell lysates. HA signal in the pull down fraction indicated retention of Fic2 by Fic1. The specificity of this interaction was confirmed by the absence of signal when partner protein Fic1 lacked the FLAG epitope. These properties indicate that Fic1 and Fic2 of *C. fetus* subsp. *venerealis* 84-112 form a functional toxin–antitoxin system. Given that the enzymatic activity of antitoxin Fic1 is autoregulated via *inh*, this protein exhibits a mode of concomitant intra- and intermolecular- toxin neutralization novel for bacterial Fic proteins.

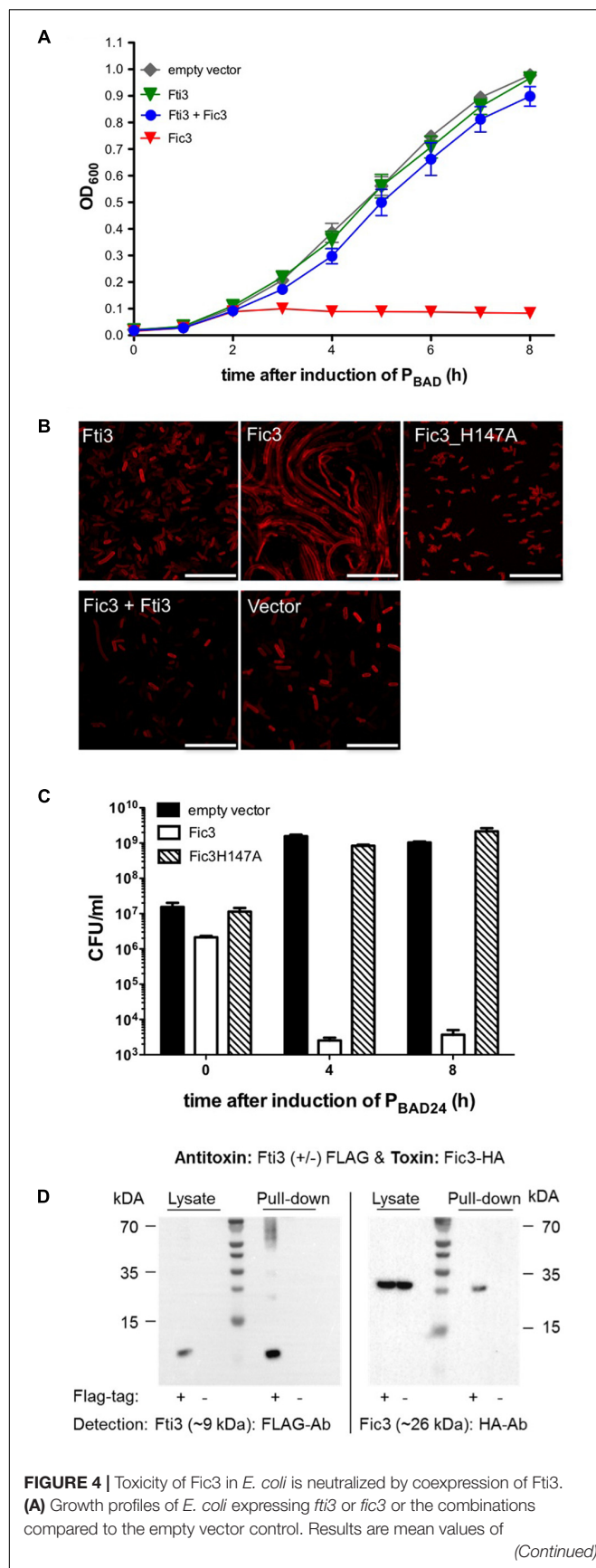
### Fti3 Acts as an Antitoxin for Fic3

Fic3, like Fic2, carries the enzyme core motif but lacks an *inh* motif (Figure 1). Similar to the result of *fic2* expression, *E. coli* carrying *fic3* failed to grow under inducing conditions (Figure 4A). However, dual expression of *fic3* and its neighboring gene encoding the putative inhibitor protein restored *E. coli* growth completely. *E. coli* cells expressing the inhibitor alone grew indistinguishably from cells carrying the vector control. Since, this protein acts as an antitoxin for Fic3 we named the gene *fti3* (Fic toxin inhibitor 3). Microscopy of the toxin/antitoxin expressing *E. coli* revealed that inhibitor protein alone had no impact on cell morphology (Figure 4B). By contrast we observed extreme filamentation due to Fic3 that could be reversed by either mutation of the catalytic histidine in variant Fic3H147A, or co-expression of wild type Fic3 with Fti3. To assess the viability of cells expressing Fic3, samples of a culture before and after 4 or 8 h of arabinose-induced expression were plated on LB agar without arabinose. Viability of the culture dropped by several orders of magnitude after Fic3 induction (Figure 4C). By contrast cells producing the non-toxic variant Fic3H147A exhibited similar viability as the vector control strain. Direct interaction between Fti3 and toxin Fic3 was tested following coexpression of FLAG-tagged Fti3 and HA-tagged Fic3. FLAG-tagged Fti3 from the cell lysate bound the affinity matrix and specifically retained Fic3-HA in the pull down reaction (Figure 4D). No retention of Fic3 was detected when Fti3 lacked the FLAG epitope. We conclude that Fti3-Fic3 form another TA module on the extrachromosomal ICE in addition to the chromosomal system *fic1-fic2*.

### Putative Antitoxin Fti4 Interacts with Fic4

Expression of *fic4* for up to 8 h had no effect on cell growth (Figures 5A,C) and affected cell morphology only mildly (Figure 5B). To test whether the protein's inhibitory motif was suppressing the predicted enzyme activity, residues Thr209 and Glu213 were exchanged for alanine. Expression of the mutant variant was compatible with normal growth comparable to cells expressing wild type Fic4 or the vector control (Figure 5A) but a filamentous phenotype was observed upon Fic4\_T209A/E213A-HA expression (Figure 5B). Since cells expressing wild type Fic4 are phenotypically normal under these conditions we used the mutant variant to test for a possible antitoxin activity for the adjacent ORE, Fti4. Coexpression of Fti4





**FIGURE 4 |** Continued

OD<sub>600</sub> ± standard deviation of three independent experiments. (B) Confocal microscopy of Nile red stained *E. coli* expressing indicated *ftl* or *fic* genes or derivatives, respectively, alone or in combination. Vector control (lower right panel), scale bars (25 μm). (C) Colonies formed (CFU per ml) 0, 4, and 8 h post-induction for *E. coli* expressing *fic3*, *fic3\_H147A* or the vector control. Results are mean values of three independent experiments.

(D) Co-immunoprecipitation of Fti3 (9 kDa) and Fic3 (26 kDa). Either FLAG-tagged (+) or native (-) Fti3 were expressed in *E. coli* together with HA-tagged Fic3 and incubated with FLAG-affinity beads. Cell lysates and elution fractions (pull-down) were loaded on gels and proteins were detected by Western analysis using indicated anti-FLAG- or anti-HA-antibodies (Ab).

and mutant Fic4 lessened filamentation substantially (Figure 5B) but had no impact on growth or survival (Figures 5A,C).

To test whether Fti4 and Fic4 physically interact, fusion proteins with epitope tags were created and simultaneously produced in *E. coli* as described above. FLAG-Fti4 (~10 kDa) was not directly detectable in the cell lysates but was visible after enrichment of the protein on the affinity matrix (Figure 5D). Fic4-HA was detected in lysates of both test and control strains. Fic4-HA was also retained on the FLAG affinity beads in a manner dependent on FLAG-Fti4. Since the functional tests described above showed phenotypes for Fti4 only when combined with the mutant derivative of Fic4, we also assayed for protein binding using the mutant allele. Similar to wild-type Fic4, co-retention of Fic4\_T209A/E213A by FLAG-Fti4 was observed (Figure 5D, right panel).

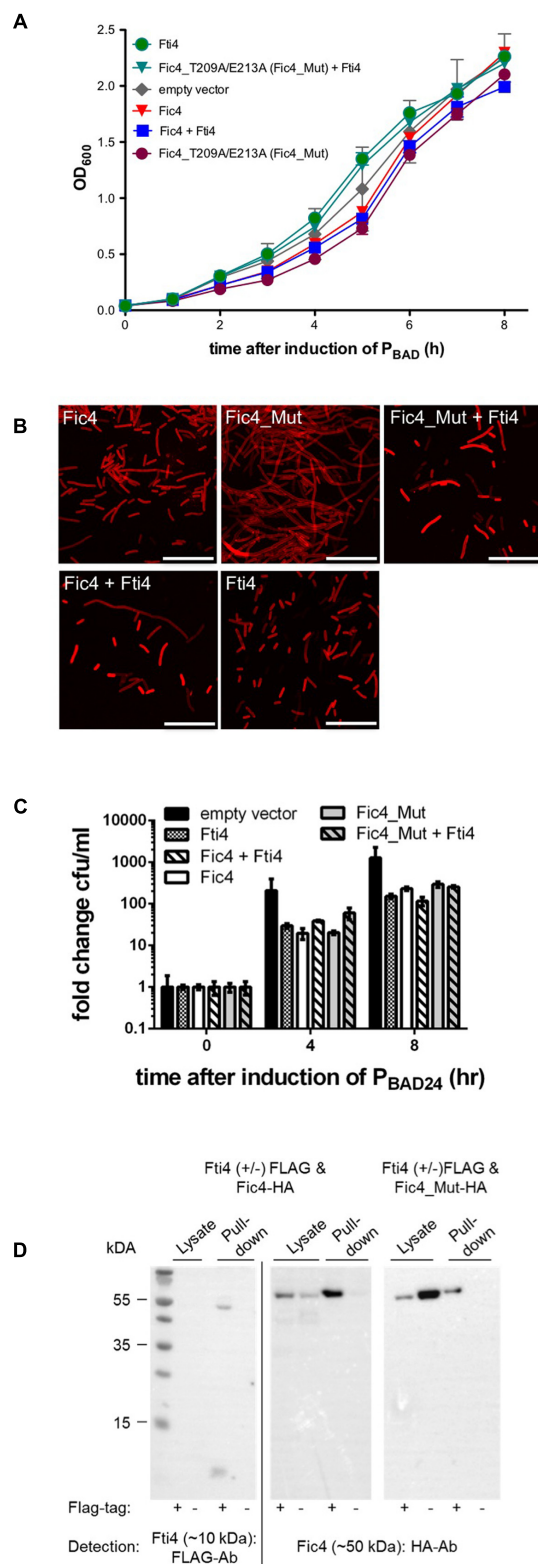
In summary, some of the observed characteristics of Fic4 are consistent with the function of a toxin, yet the toxicity of the mutant form was quite mild compared to the inh-deficient Fic1 derivative and the wild type class I proteins Fic2 and Fic3. It is possible that evolution has introduced mutations outside of the Fic4 active site that impair enzyme activity. It is further possible that the surrogate host *E. coli* simply lacks the specific protein targeted by Fic4. Another hypothesis that we could test was to ask whether Fic4 might actually function as an antitoxin for a distinct locus (below).

## Fic2 Toxicity Is Inactivated *in Trans*

To explore potential *in trans* interactions involving components of the distinct systems each toxin was expressed pairwise with every putative antitoxin. We found that the ICE<sub>84-112</sub> encoded antitoxin Fti3 reversed the growth defect caused by toxin Fic2 (Figure 6A). In contrast co-expression of Fic4 with Fic2 had no effect. Neutralization of Fic2 toxicity by Fti3 was confirmed by plating samples of the induced cultures (Figure 6B). Cells survived 4 and 8 h of Fic2 expression when co-expressing Fti3, but Fic4 was not able to counteract the toxicity of Fic2. Functional interaction between Fic2 and Fti3 was also supported by the normalized morphology of cells following co-expression compared to the filamentous phenotype caused by Fic2 alone (Figure 6C). A partial reversal of the Fic2-induced filamentation was observed with Fic4.

To test for direct binding interactions between the protein pairs, pull-down assays were performed with cells expressing Fic2-HA with either FLAG-tagged Fti3 or FLAG-tagged Fic4.





**FIGURE 5 |** Fti4 interacts with the non-toxic Fic4. **(A)** Growth profiles of *E. coli* expressing indicated *fti* and *fic* genes, or derivatives, alone or in combination. Results are mean values of OD<sub>600</sub> ± standard deviation of three (Continued)

**FIGURE 5 |** Continued

independent experiments. **(B)** Confocal microscopy of Nile red stained cells cultured in **(A)** as indicated. Fic4\_Mut refers to Fic4\_T209A/E213A. Scale bars (25 μm). **(C)** Colonies formed (CFU per ml) 0, 4, and 8 h post-induction for *E. coli* expressing *fic4* and *fic4\_209A/E213A* with or without Fti4, or vector control. **(D)** Co-immunoprecipitation of Fti4 (10 kDa) and Fic4 (50 kDa). Either FLAG-tagged (+) or native (-) Fti4 were expressed in *E. coli* together with HA-tagged Fic4 and incubated with FLAG-affinity beads. Cell lysates and elution fractions (pull-down) were loaded on gels and proteins were detected by Western analysis using indicated anti-FLAG- or anti-HA-antibodies (Ab).

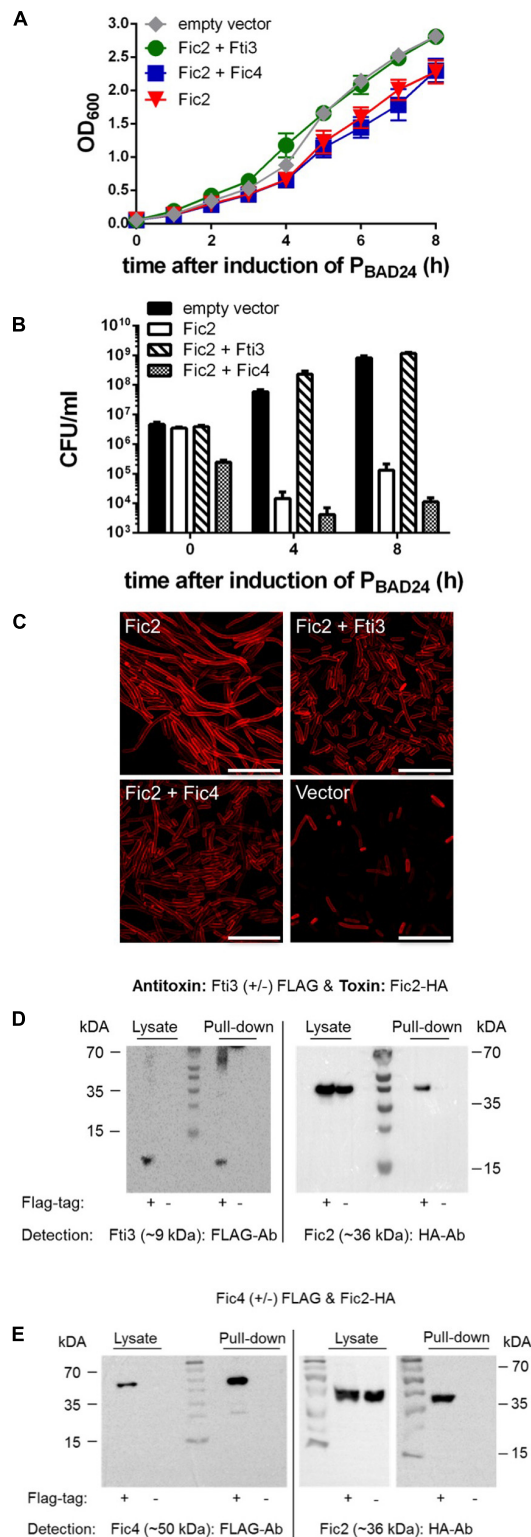
Consistent with the functional results shown above, Fti3 was able to retain Fic2 (**Figure 6D**). Remarkably, although Fic4 showed little antitoxin activity for Fic2, a complex of these proteins was detected nonetheless (**Figure 6E**).

To complete the analyses for toxin Fic2, the same tests were performed for the last putative antitoxin Fti4. No neutralizing activity by Fti4 was observed during cellular growth or by monitoring cell morphology. The pull down assay combining FLAG-tagged Fti4 with Fic2-HA also failed to detect interaction between these proteins (all data not shown). Given that Fti4 is 48% identical to Fti3, the lack of activity observed for Fti4 shows that the ability of antitoxin Fti3 to inactivate Fic2 is specific. As a final specificity check we also tested whether the toxic form of Fic1 (Fic1\_S31A/E35A) was affected by co-production of either Fti3 or Fti4. No reversion of the poor growth, reduced survival or filamentous phenotypes caused by Fic1\_S31A/E35A was observed (data not shown).

The sum of these data demonstrate that the bacterial growth phenotype caused by Fic2 is counteracted by the *cis* encoded antitoxin, Fic1, and independently by Fti3 *in trans*. Fic4 partially reversed the toxic effect of Fic2 in *E. coli*. Both the *cis* acting antitoxin Fic1 and the ICE-encoded proteins Fti3 and Fic4 were shown to bind toxin Fic2. These findings support a model of functional crosstalk occurring between chromosomally and ICE\_84-112 encoded Fic proteins that act to control the toxin Fic2.

## Fic4 Interacts with Toxin Fic3

To test for potential regulatory crosstalk occurring between Fic3 and antitoxins of the distinct systems, we again performed phenotypic tests following dual expression of each protein pair. Both Fic1 and Fic4 were unable to neutralize the extreme growth phenotype caused by Fic3 (**Figure 7A**). Dual expression of Fic3 with Fic1 did not revert the filamentation induced by Fic3, but partial recovery was apparent upon co-expression of Fic3 with Fic4, suggesting some neutralizing interactions (**Figure 7B**). Consistent with these results the pull-down assay was clearly negative for binding between Fic1 and Fic3 (**Figure 7C**), but a small yield of co-purified Fic3 was detected using FLAG-tagged Fic4 (**Figure 7D**). We performed the same analyses with cells coexpressing Fti4 with Fic3. Again despite its similarity to antitoxin Fti3, Fti4 had no affect on Fic3 toxicity and the proteins failed to bind under these conditions (not shown). In summary, we conclude that Fic3 is effectively neutralized by the *cis* encoded Fti3. Moreover, modest levels of complex formation



**FIGURE 6 |** Fic2 interacts with Fti3 and Fic4 and Fic2 toxicity is neutralized by Fti3. Growth profiles over time (A) and corresponding colonies formed (CFU per ml) (B) at 0, 4, and 8 h post-induction of *E. coli* expressing *fic2* alone or in combination with either *fti3* or *fic4*. Results are mean values of OD<sub>600</sub> or (Continued)

**FIGURE 6 |** Continued

CFU/ml  $\pm$  standard deviation of three independent experiments. (C) Confocal microscopy of Nile red stained *E. coli* expressing indicated *fic2* alone or in combination with *fti3* or *fic4*. Vector control (lower right panel), scale bars (25  $\mu$ m). (D) Co-immunoprecipitation of Fti3 (9 kDa) and Fic2 (36 kDa) and (E) Fic4 (50 kDa) and Fic2 (36 kDa). Either FLAG-tagged (+) or native (-) Fti3 (D) or Fic4  $\pm$  FLAG (E) were expressed in *E. coli* together with HA-tagged Fic2 and incubated with FLAG-affinity beads. Cell lysates and elution fractions (pull-down) were loaded on gels and proteins were detected by Western analysis using indicated anti-FLAG- or anti-HA-antibodies (Ab).

with trans-acting factor Fic4 may contribute to regulation of this enzyme.

## Fic2 or Fic3 Expression Inhibits Translation in *E. coli*

The identity of specific protein targets modified by Fic enzymes in bacteria is difficult to predict. It is known, however, that the activities of many TA toxins interfere with the translation process either directly, e.g., by cleavage of mRNA or tRNA, or as a downstream effect (Rajashekar et al., 2009; Park et al., 2013). To measure translation in *E. coli* cells expressing *C. fetus* Fic proteins, we performed sucrose gradient centrifugation of cell lysates and recorded polysome profiles (Figures 8A,B). In these analyses, the height of the polysome peaks is directly proportional to the translation levels, therefore translation defects can be faithfully detected by the reduction of polysome peaks. Moreover, because free ribosomal subunits, 70S monosomes and polysomes can be resolved, changes in the ratios between these different ribosomal (sub-) complexes can give additional information on the type of defect causing reduced translation. We recorded profiles from *fic*-expressing *E. coli* cells and compared them to those of the vector control strain. The signal corresponding to ribosomal subunits and translating ribosomes is indicated for each gradient. To make *fic*-dependent shifts in the relative abundance of these populations more apparent, profiles from different expressing strains were overlaid in the figure. Expression of Fic3 inhibited translation severely, as obvious from the massive reduction of polysome levels (Figure 8A, red trace; Table 1) compared to the vector control strain (black trace) or cells expressing just antitoxin Fti3 (green trace). Concomitantly, a strong increase of the 70S peak was detected, suggesting that ribosomal subunits are competent for joining into 70S ribosomes, but fail to enter into translation. We conclude that Fic3 blocks a step after subunit joining but before translation elongation. Dual expression of Fic3 and antitoxin Fti3 (blue trace) largely restored translation to normal levels. Moreover, the abnormally high 70S peak observed upon Fic3 overexpression was partially reduced upon Fti3 co-expression.

Fic2 expression also had a mild inhibitory effect on translation, as obvious from an accumulation of free 30S and 50S ribosomal subunits relative to the amount of 70S ribosomes and polysomes (Figure 8B). We compared the free subunit accumulation relative to the polysome abundance in multiple independent experiments ( $n = 7$ ) and determined a quantitatively significant increase (Table 1). Notably, in contrast to Fic3 expression, Fic2 expression

did not result in increased 70S levels, and even reduced 70S amounts compared to the vector control. Reduced 70S and increased free subunit levels are indicative of inefficient subunit joining. *E. coli* cells expressing antitoxin Fic1 (green trace) showed no significant variation in the ratio of free subunits versus polysomes compared to profiles from the vector control strain (Figure 8B and Table 1). The relative abundance of 70S species was mildly reduced in *fic1*-expressing vs. vector control cells however (Table 1), consistent with the observation that Fic1 expression slightly inhibits cell growth (Figure 3). In line with the neutralizing activity observed in our previous functional tests, simultaneous expression of antitoxin Fic1 with Fic2 (blue trace) resulted in a profile similar to the empty vector control.

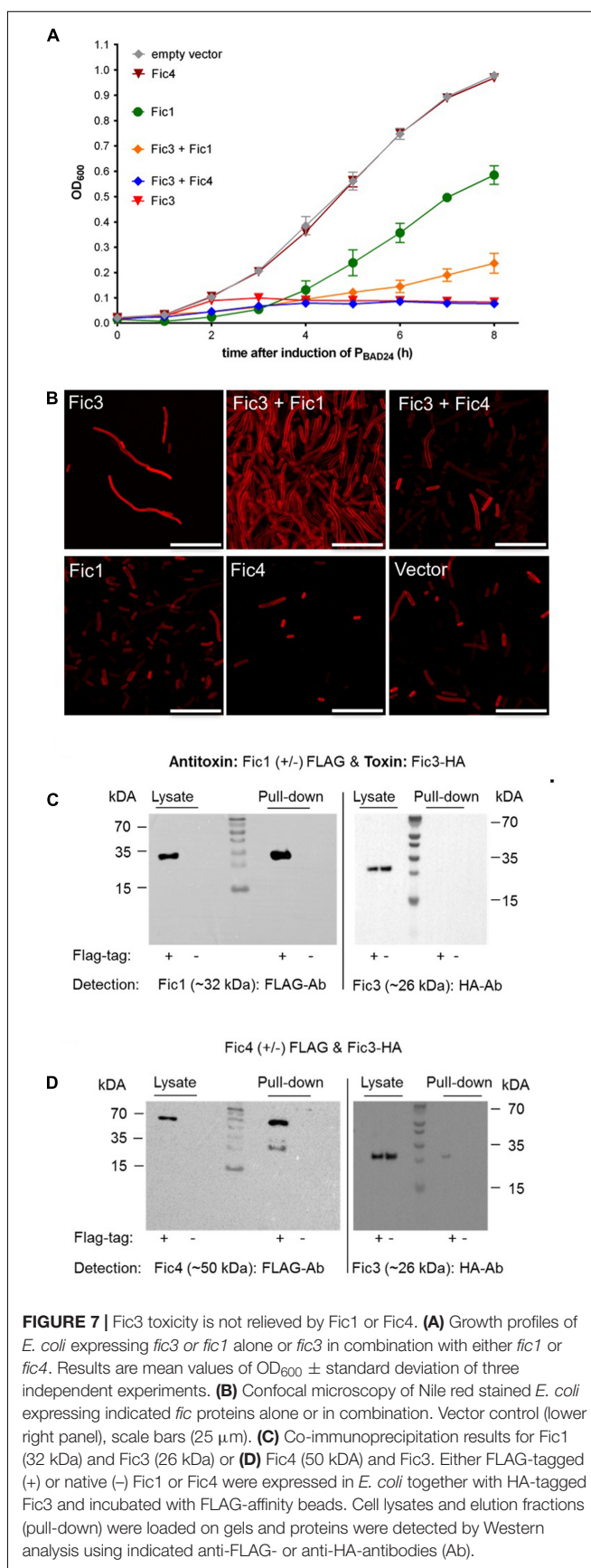
## Prevalence of *fic* Genes within *C. fetus* Strains

The sum of our findings suggests that the *C. fetus* *fic* genes act as TA systems. In that case the loci should be well-conserved within the species. We used PCR to survey the prevalence of the *fic* genes and *fti3* in 102 *C. fetus* isolates from geographically and ecologically diverse sources (summary in Table 2; detailed information in Supplementary Table S5). All of the *C. fetus* subsp. *venerealis* strains ( $n = 62$ ) were positive for *fic1* and 59 out of 62 (95%) were positive for *fic2*. This finding is consistent with genetic linkage of the Fic2 toxin to the Fic1 antitoxin. Sequence analysis of full-length *fic2* amplicons randomly selected from our strain collection ( $n = 6$ ) showed complete conservation for this subspecies (data not shown). In contrast, only 5 out of 40 (12.5%) *C. fetus* subsp. *fetus* strains harbor *fic1* and only two carry the *fic2* gene, whereby strain *C. fetus* subsp. *fetus* 98/v445 (F37) lacks the corresponding *fic1* antitoxin gene. Sequence analysis of this solitary *fic2* allele revealed 36 nucleotide changes, corresponding to 13 amino acid substitutions. Expression of the F37 *fic2* gene in *E. coli* confirmed that the mutated toxin is functionally impaired (data not shown). The *fic3* and *fic4* genes were detected exclusively in *C. fetus* subsp. *venerealis*, *fic3* in 11.3% (7/62) and *fic4* in 4.8% (3/62) of the isolates. Gene *fti3* shows higher abundance: 59.7% (37 of 62) *C. fetus* subsp. *venerealis* isolates and four *C. fetus* subsp. *fetus* isolates (4/40) carry the gene. Consistent with the predicted selective pressure for co-existence, all strains positive for *fic3* additionally encode the corresponding antitoxin Fti3. Moreover the high prevalence of *fic2* in *C. fetus* subsp. *venerealis* may select for stable maintenance of *fti3* even in the absence of the cognate toxin *fic3*.

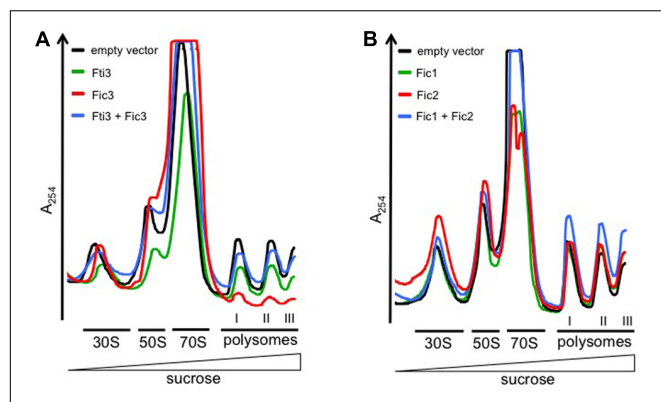
In summary, we conclude that the presence of a *fic* toxin gene in *C. fetus* is typically linked to carriage of the paired antitoxin gene. The chromosomal TA system is highly conserved in *C. fetus* subsp. *venerealis*. The ICE-associated loci are also unique for *C. fetus* subsp. *venerealis* but are comparatively rare in the strains surveyed.

## Phylogenetic Analysis

The significant association of these *fic* genes with *C. fetus* subsp. *venerealis* and their relative absence in *C. fetus* subsp. *fetus* led us to next ask whether they are present in other *Campylobacter*







**FIGURE 8 |** Fic2 and Fic3 disrupt the translational machinery. Comparisons of ribosomal profiles of *E. coli* expressing (A) *fti3*, *fic3*, the combination (*fti3* + *fic3*) compared to the vector control, or (B) *fic1*, *fic2*, or the combination (*fic1* + *fic2*) compared to the vector control. Absorption peaks corresponding to free 30S and 50S ribosomal subunits, 70S ribosomes, and polysomes are designated. The deficit in polysomes relative to free 30S subunits provoked by *fic2* and *fic3* expression is quantified in **Table 1**.

species and/or whether they are conserved in bacteria which inhabit the urogenital tract. Using the HPFXXGNXR motif in a BlastP analysis revealed that genomes of several *Campylobacter*

**TABLE 1 |** Numerical and statistical analysis of ribosomal subunit peaks.

		Ratio subunits vs. polysome (I)		
		30S	50S	70S*
<sup>A</sup> Set 1	Vector	1.06	1.42	6.76
	Fic1	1.08	1.52	3.15
	Fic2	1.47	1.70	2.64
	Fic1 + Fic2	0.86	1.26	3.07
<sup>B</sup> Set 2	Vector	1.06	1.40	3.21
	Fti3	1.16	1.34	3.38
	Fic3	1.94	nd	11.84
	Fti3 + Fic3	1.10	1.67	1.67
		Vector	Fic2	<i>p</i> -Values**
Statistic ( <i>n</i> = 7)	30S	1.10 ± 0.02	1.50 ± 0.19	0.0005
	50S	1.49 ± 0.11	2.03 ± 0.38	0.0137

<sup>A,B</sup>see **Figure 8**; Numbers represent ratios of 30S or 50S subunit versus first polysome peak (I) values. Numbers were calculated based on XSpan generated "Heights" (see section "Materials and Methods"); nd, 50S peaks are not separated enough from 70S peaks for numerical analysis with XSpan; \*Height was generated by XSpan after extrapolation of clipped curves. \*\**p*-Values were calculated from seven profiles using the paired Student's *t*-test.

**TABLE 2 |** Prevalence of *fic* genes in *C. fetus* subspecies.

		Chromosomal			ICE_84-112			
	#	<i>fic1</i>	<i>fic2</i>	<i>fic1+fic2</i>	<i>fic3</i>	<i>fic4</i>	<i>fti3</i>	<i>fti3+fic3</i>
<i>Cff</i> <sup>a</sup>	40	5	2	1	0	0	4	0
<i>Cfv</i> <sup>a</sup>	62	62	59	59	7	3	37	7

<sup>a</sup>*Cff* *C. fetus* subsp. *fetus*, *Cfv* *C. fetus* subsp. *Venerialis*.

species encode from one to four Fido proteins. BlastP analyses were then performed with the full-length *C. fetus* proteins Fic1-4 to identify related Fido proteins from epsilon-proteobacteria or distantly related bacteria. Interestingly, BlastP analysis using full-length Fic2 or the degenerated motif of Fic2 consequently retrieved proteins of bacterial species linked to human fertility complications (Moreno et al., 2016; Pelzer et al., 2017). We used these proteins, related *Campylobacter* proteins and selected reference Fic-proteins (Harms et al., 2016b) to generate the Neighbor joining tree shown in **Figure 9**. The tree architecture placed the proteins in two main branches. Cluster A includes Fic3 and Fic4 of *C. fetus* subsp. *venerialis* 84-112. Reference FIC-domain proteins of *Bartonella*, *Yersinia enterocolitica*, and *E. coli* (marked with asterisks) were also placed in this cluster.

Fic1 and Fic2 are both grouped in the B branch and resolve in separate subclusters. Interestingly, *Fusobacterium* spp. also harbor FIC proteins closely related to Fic1. *Fusobacteria* inhabit mucous membranes of humans and animals and both *Fusobacterium nucleatum* and *Fusobacterium necrophorum* cause abortion in cattle (Kirkbride et al., 1989; Otter, 1996). Moreover *F. nucleatum* causes intra-amniotic infection and premature delivery in humans and mice (Han et al., 2004; Gauthier et al., 2011). We also note that *C. ureolyticus* ACS-301-V-Sch3b, isolated from the human vaginal tract, encodes a protein related to Fic1 and another related to Fic3 and Fic4 of cluster A.

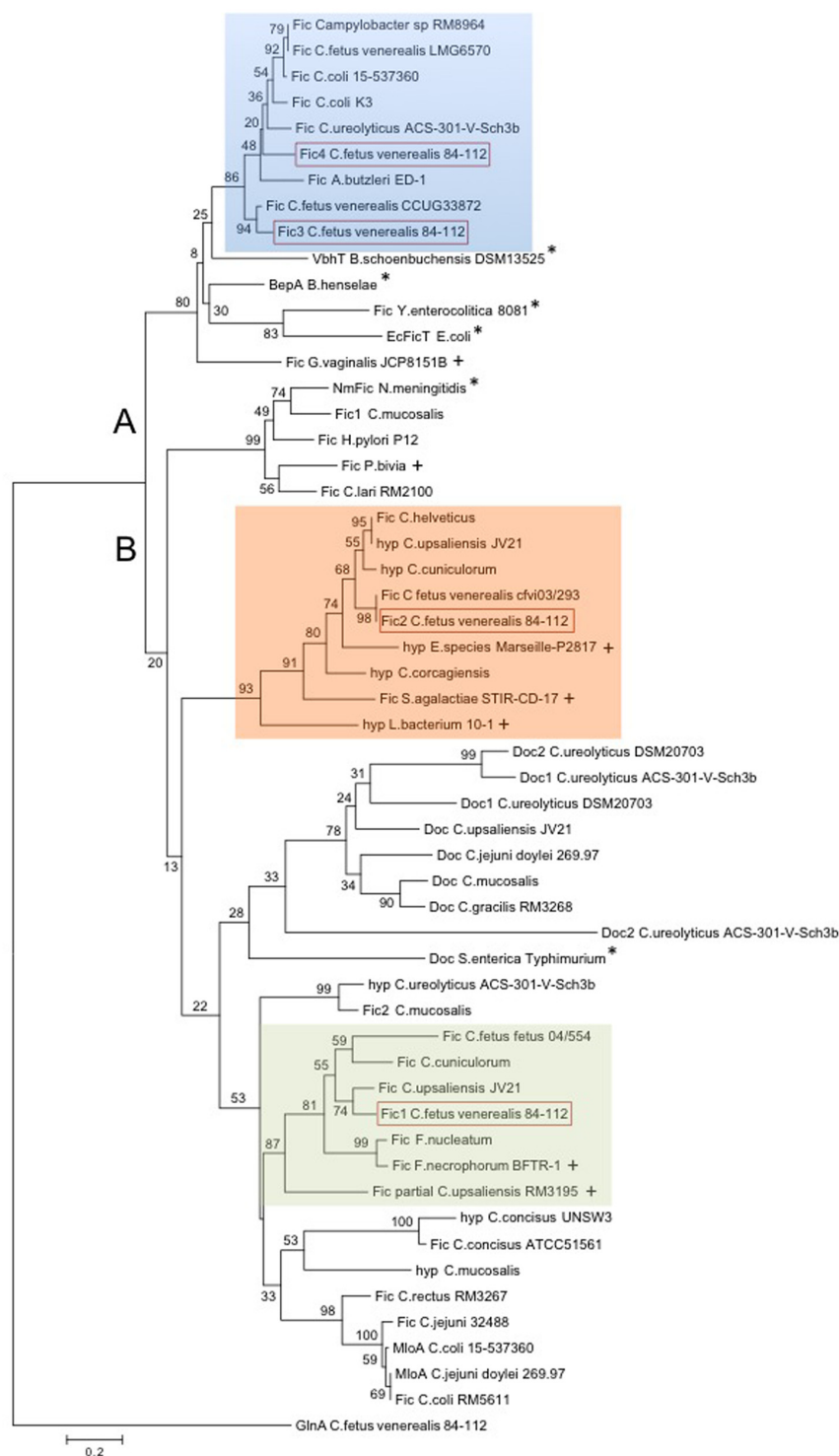
Fic2 clusters with a hypothetical protein of *Campylobacter upsaliensis* JV21 and proteins of *C. helveticus* and the more recently described novel species *C. cuniculorum* and *C. corgagiensis*. *C. upsaliensis* is a human enteropathogen that typically causes diarrhea, bacteremia and sepsis, but human infection with *C. upsaliensis* has also been associated with spontaneous abortion (Gurgan and Diker, 1994). Moreover, proteins of bacterial species suspected or confirmed to play a role in human fertility and pregnancy outcome cluster in the same branch (indicated with a plus sign).

In conclusion, our findings show conserved FIC protein sequences in a variety of bacteria that either inhabit the urogenital tract of humans and animals or are able to establish pathology in this niche.

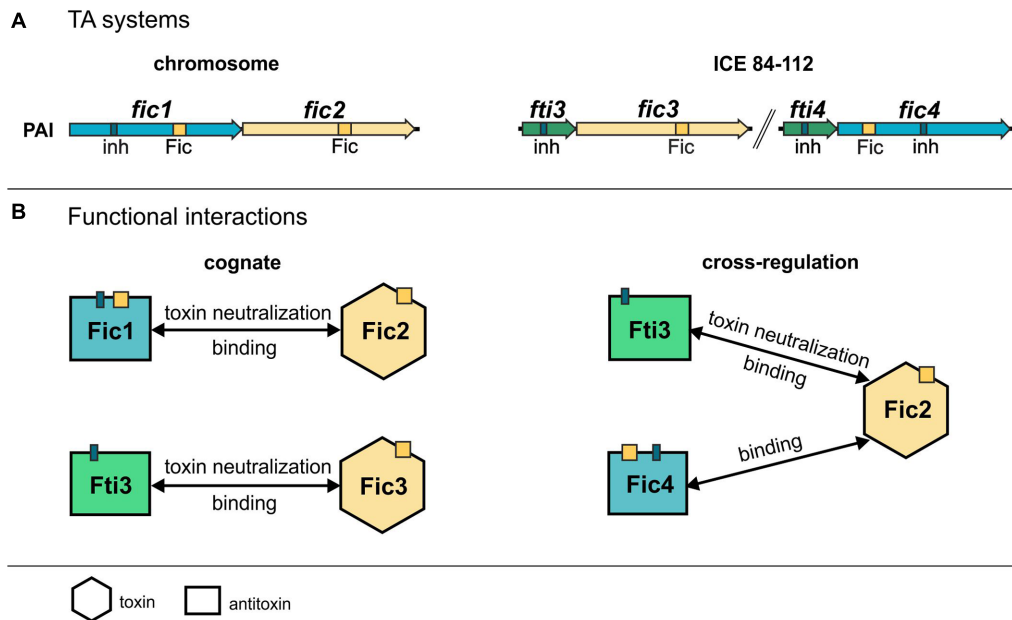
## DISCUSSION

*Campylobacter fetus* subsp. *venerialis* 84-112 expresses a group of Fic proteins that are conserved in various veterinary and human pathogens of the urogenital tract. This study shows that the *fic* modules are prevalent and strongly conserved in *C. fetus* subsp. *venerialis* isolates but generally lacking in *C. fetus* subsp. *fetus*. The data also provide the first experimentally validated examples of TA activity for Fic proteins in *Campylobacter*. Bacterial genomes generally harbor multiple TA modules and it is becoming increasingly clear that the TA-associated toxins perform discrete, multipurpose functions (Ramage et al., 2009; Leplae et al., 2011; Lobato-Marquez et al., 2016; Diaz-Orejas et al., 2017). *C. fetus* subsp. *venerialis* 84-112 carries one functional module on the chromosome. Two additional systems are located on the extra chromosomal element





**FIGURE 9 |** Neighbor joining tree showing phylogenetic relationship of Fido proteins in *Campylobacter* spp. BlastP search of the conserved motif HPFXXGNXR, Fic2 motif HPFREGNTRTIA, and full-length Fic1-Fic4 proteins of *C. fetus* subsp. *venerealis* 84-112 revealed Fido proteins in *Campylobacter* spp., in closely related epsilon-proteobacteria, including *Arcobacter* (*A. butzleri*), *Helicobacter* (*H. pylori*), as well as in more distant species, as indicated. The model template for protein structure prediction of Fic4, VbhT of *B. schoenbuchensis*, as well as other well-described reference Fic proteins were included (asterisk). The Neighbor joining tree contains 54 proteins and is rooted to the *C. fetus* subsp. *venerealis* 84-112 housekeeping protein GlnA. Protein and organism names are shown. Fic1-4 of *C. fetus* subsp. *venerealis* 84-112 are highlighted (red box). Proteins from distant species linked to infertility or abortion are indicated with pluses. Protein accession numbers are listed in Supplementary Table S5. The two obtained clusters (A and B) are indicated. Bootstrap values (1,000 replicates) are shown at the tree nodes. The scale bar represents 0.2 substitutions per amino acid position.



**FIGURE 10 |** Summary of cognate and non-cognate rescue and interaction data. **(A)** TA modules of *C. fetus* subsp *venerealis* strain 84-112 are depicted. The chromosomal locus encodes AT-Fic1/T-Fic2. The extrachromosomal ICE carries two additional independent modules AT-Fti3/T-Fic3 and the module encoding the putative TA system AT-Fti4/T-Fic4. Conserved active site motifs (Fic) and inhibitory helices (inh) are indicated. **(B)** The functional interaction and physical binding of cognate TA components detected in this study are shown (left). Fti4 binds Fic4 (not shown), but Fic4 has no toxic activity in *E. coli*. Cross-regulation of toxin Fic2 by non-cognate components is shown (right). Two unrelated ICE-encoded proteins, Fic4 and Fti3, bind T-Fic2. Heterologous AT-Fti3 counteracted Fic2 toxicity.

ICE\_84-112 (Kienesberger et al., 2014). It has been speculated that TA components encoded by horizontally acquired DNA and chromosomal loci evolve toward functional cooperation (Saavedra De Bast et al., 2008; Makarova et al., 2009). Here, we demonstrate that Fti3 encoded by ICE\_84-112 provides immunity for the chromosomal toxin Fic2 in addition to the cognate antitoxin Fic1 (see summary of results, **Figure 10**). We also show that the ICE-encoded Fic4 binds Fic2 and may therefore contribute to toxin regulation by complex formation or sequestration. Oligomerization was shown to play a role in regulating toxin activity for the class III Fic protein NmFic from *Neisseria meningitidis* (Stanger et al., 2016). In that case, activation of the NmFic toxin is blocked by tetramer formation. Control of *C. fetus* toxin Fic2 by Fic1 and possibly Fic4 implies that inhibitory strategies involving heteromeric complexes of different toxins are also possible. The structural similarities of the *C. fetus* Fic proteins may support physical interactions between non-cognate components as shown with the *ccd* and *parD* systems (Smith et al., 2012). Compared to well-studied Fic proteins or paradigm type II TA systems evidence for physical and functional interactions between non-cognate toxin-antitoxin systems is still relatively rare but clearly emerging as shown for *Mycobacterium tuberculosis* (Yang et al., 2010; Zhu et al., 2010). TA systems can also be interconnected through transcriptional regulation, for example positive feedback regulation can allow production of toxins to induce transcription of other TA systems (Kasari et al., 2013). As another example, toxin MqsR of the type II TA system MqsR/MqsA regulates GhoT toxin of a type V

TA system via post-transcriptional differential mRNA cleavage. This activity results in a regulatory hierarchy where one TA system controls another (Wang et al., 2013). Wessner et al. (2015) have also established that regulatory crosstalk occurs between modules of type-I and -II TA families in the human pathogen *Enterococcus faecalis*. Taken together these data support the notion that bacteria harboring multiple TA systems may develop a complex hierarchy. Cooperative regulation of their activities would support concerted physiological responses in the cell. Moreover, interplay between TAs may enable bacteria to create heterogeneous populations and survive stress under a wider range of environmental conditions (Fasani and Savageau, 2013).

The ability to enter prolonged dormancy is an important factor in the epidemiology and spread of *Campylobacter* (Rollins and Colwell, 1986; Bronowski et al., 2014). Dormancy requires the bacterial cells to switch from normal activities to a static state (Harms et al., 2016a). Enzymes belonging to the Fido family are well-suited to control this process as they can modify a broad range of cellular proteins post-translationally (Garcia-Pino et al., 2014; Roy and Cherfils, 2015; Harms et al., 2016b). The general lack of similarity outside of the active loop implies that Fic1-4 bind to distinct protein targets. The core motif conserved in Fic1, Fic3, and Fic4 suggests these are competent for adenylation. The modification reaction catalyzed by Fic2 remains unclear because the core motif deviates from the adenylation consensus. Regardless of the biochemistry involved, we found that expression of both the canonical

Fic3 and the degenerate Fic2 toxins in *E. coli* interferes with translation. The inhibitory effect of Fic3 on translation was very pronounced. Fic3 expressing cells show a very high 70S peak and drastic reduction of polysome levels suggesting that ribosomal subunits are produced and can join, but are incapable of entering into translation. In contrast, Fic2 reduced 70S levels, while the amounts of free ribosomal subunits were increased. These properties are consistent with an early defect that impairs subunit joining. We conclude that Fic2 either directly affects subunit joining, or alternatively, that it inhibits a step of ribosome biogenesis, thereby causing the synthesis of aberrant, joining defective ribosomal subunits. Fic2 might therefore modify an rRNA processing factor or an assembly cofactor. We note with interest that ribosome profiles of cells overexpressing YfjG (RatA), a toxin of the *yfjG-yfjF* operon on the *E. coli* chromosome are similar to those expressing Fic2. YfjG inhibits 70S ribosome association and blocks the translation initiation step (Zhang and Inouye, 2011). Work in other laboratories has shown that mutation or depletion of ribosome assembly GTPases, but also inhibition of translation, is associated with filamentous cell morphology (Karbstein, 2007). These attributes resemble the phenotypes we observed. To better understand the mechanism of bacterial cytotoxicity we are characterizing each of the *C. fetus* subsp. *venerealis* toxins biochemically and structurally.

Although we can generally conclude that the TA systems described here exist in *C. fetus* subsp. *venerealis* 84-112 to control the switch between normal and static metabolic states, details about the biological context of that activity remain unknown. Emerging data from animal models of uropathogenic *E. coli* infection establish that TA systems are important for niche-specific colonization and survival, and a contribution to virulence was described in *Salmonella* Typhimurium (Norton and Mulvey, 2012; De la Cruz et al., 2013; Lobato-Marquez et al., 2015). Thus, presence of multiple Fic proteins in *C. fetus* subsp. *venerealis* may enhance long term survival under hostile conditions within the host or in response to stress during its environment-animal host infectious cycle (Man, 2011). Fasani and Savageau (2013) have proposed a general model of TA systems in which redundancy of the systems is important for increasing the frequency of persister cells. Evidence is further emerging that TA modules can contribute directly to the virulence repertoire of bacteria (Lobato-Marquez et al., 2016). Conservation of related Fic proteins in

isolates of *Arcobacter*, *Bartonella*, *Fusobacterium*, *Streptococcus*, *Lachnospiraceae*, *Prevotella*, *Gardnerella*, and *Enterococcus* that colonize or cause disease in urogenital or feto-placental tissue in humans and livestock underscores the probable importance of this group of Fic proteins for niche adaptation and pathogenicity. Characterizing the protein-interaction networks of the Fic proteins of *C. fetus* and analogs from other urogenital pathogens will be the next step in understanding these complex multipurpose toxins.

## AUTHOR CONTRIBUTIONS

HS, SK, GG, and EZ designed the research. CH contributed study resources. HS, SK, BP, LP, BK, PB, DV, and DA performed experiments. HS, SK, BP, FF, and EZ analyzed the data. HS, SK, BP, GG, and EZ wrote the paper. All authors read and approved the final manuscript.

## FUNDING

This study was supported by the Austrian Science Fund FWF grants P20479 (GG and EZ) P24016 (EZ), and the DK Molecular Enzymology W901 (EZ), BioTechMed-Graz, NAWI-Graz (EZ), the funds of the Oesterreichische Nationalbank (Anniversary Funds, project number: 14321 to CH), and the Hygiene Fund Young Scientist grant from the Medical University of Graz (SK).

## ACKNOWLEDGMENTS

We thank S. Raffl for technical assistance, K. Gruber for generating the protein structure models and H. Wolinski and K. Hellauer for expert assistance in microscopy.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.01965/full#supplementary-material>

## REFERENCES

- Blaser, M. J., Newell, D. G., Thompson, S. A., and Zechner, E. L. (2008). "Pathogenesis of *Campylobacter fetus* infections," in *Campylobacter*, 3rd Edn, eds I. Nachamkin, C. M. Szymanski, and M. J. Blaser (Washington, DC: American Society for Microbiology), 401–428.
- Bronowski, C., James, C. E., and Winstanley, C. (2014). Role of environmental survival in transmission of *Campylobacter jejuni*. *FEMS Microbiol. Lett.* 356, 8–19. doi: 10.1111/1574-6968.12488
- Bullman, S., Lucid, A., Corcoran, D., Sleator, R. D., and Lucey, B. (2013). Genomic investigation into strain heterogeneity and pathogenic potential of the emerging gastrointestinal pathogen *Campylobacter ureolyticus*. *PLOS ONE* 8:e71515. doi: 10.1371/journal.pone.0071515
- Castro-Roa, D., Garcia-Pino, A., De Gieter, S., van Nuland, N. A., Loris, R., and Zenkin, N. (2013). The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat. Chem. Biol.* 9, 811–817. doi: 10.1038/nchembio.1364
- Christie, P. J., Whitaker, N., and Gonzalez-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta* 1843, 1578–1591. doi: 10.1016/j.bbamcr.2013.12.019
- Cruz, J. W., Rothenbacher, F. P., Maehigashi, T., Lane, W. S., Dunham, C. M., and Woychik, N. A. (2014). Doc toxin is a kinase that inactivates elongation factor Tu. *J. Biol. Chem.* 289, 7788–7798. doi: 10.1074/jbc.M113.544429
- Cruz, J. W., and Woychik, N. A. (2014). Teaching Fido new ModiFICATION tricks. *PLOS Pathog.* 10:e1004349. doi: 10.1371/journal.ppat.1004349

- De la Cruz, M. A., Zhao, W., Farenc, C., Gimenez, G., Raoult, D., Cambillau, C., et al. (2013). A toxin-antitoxin module of *Salmonella* promotes virulence in mice. *PLOS Pathog.* 9:e1003827. doi: 10.1371/journal.ppat.1003827
- Diaz-Orejas, R., Espinosa, M., and Yeo, C. C. (2017). The importance of the expendable: toxin-antitoxin genes in plasmids and chromosomes. *Front. Microbiol.* 8:1479. doi: 10.3389/fmicb.2017.01479
- Engel, P., Goepfert, A., Stanger, F. V., Harms, A., Schmidt, A., Schirmer, T., et al. (2012). Adenylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature* 482, 107–110. doi: 10.1038/nature10729
- Fasani, R. A., and Savageau, M. A. (2013). Molecular mechanisms of multiple toxin-antitoxin systems are coordinated to govern the persister phenotype. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2528–E2537. doi: 10.1073/pnas.1301023110
- Garcia-Pino, A., Christensen-Dalsgaard, M., Wyns, L., Yarmolinsky, M., Magnuson, R. D., Gerdes, K., et al. (2008). Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. *J. Biol. Chem.* 283, 30821–30827. doi: 10.1074/jbc.M805654200
- Garcia-Pino, A., Zenkin, N., and Loris, R. (2014). The many faces of Fic: structural and functional aspects of Fic enzymes. *Trends Biochem. Sci.* 39, 121–129. doi: 10.1016/j.tibs.2014.01.001
- Gauthier, S., Tetu, A., Himaya, E., Morand, M., Chandad, F., Rallu, F., et al. (2011). The origin of *Fusobacterium nucleatum* involved in intra-amniotic infection and preterm birth. *J. Matern. Fetal Neonatal Med.* 24, 1329–1332. doi: 10.3109/14767058.2010.550977
- Goeders, N., and Van Melder, L. (2014). Toxin-antitoxin systems as multilevel interaction systems. *Toxins* 6, 304–324. doi: 10.3390/toxins6010304
- Goepfert, A., Stanger, F. V., Dehio, C., and Schirmer, T. (2013). Conserved inhibitory mechanism and competent ATP binding mode for adenylyltransferases with Fic fold. *PLOS ONE* 8:e64901. doi: 10.1371/journal.pone.0064901
- Gorkiewicz, G., Kienesberger, S., Schober, C., Scheicher, S. R., Gully, C., Zechner, R., et al. (2010). A genomic island defines subspecies-specific virulence features of the host-adapted pathogen *Campylobacter fetus* subsp. *venerealis*. *J. Bacteriol.* 192, 502–517. doi: 10.1128/JB.00803-09
- Graaf-van Bloois, L., Miller, W. G., Yee, E., Gorkiewicz, G., Forbes, K. J., Zomer, A. L., et al. (2016). *Campylobacter fetus* subspecies contain conserved type IV secretion systems on multiple genomic islands and plasmids. *PLOS ONE* 11:e0152832. doi: 10.1371/journal.pone.0152832
- Gruber, C. J., Lang, S., Rajendra, V. K., Nuk, M., Raffl, S., Schilbach, J. F., et al. (2016). Conjugative DNA transfer is enhanced by plasmid R1 partitioning proteins. *Front. Mol. Biosci.* 3:32. doi: 10.3389/fmolb.2016.00032
- Gurgan, T., and Diker, K. S. (1994). Abortion associated with *Campylobacter upsaliensis*. *J. Clin. Microbiol.* 32, 3093–3094.
- Han, Y. W., Redline, R. W., Li, M., Yin, L., Hill, G. B., and McCormick, T. S. (2004). *Fusobacterium nucleatum* induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. *Infect. Immun.* 72, 2272–2279. doi: 10.1128/IAI.72.4.2272-2279.2004
- Harms, A., Maisonneuve, E., and Gerdes, K. (2016a). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* 354, aaf4268.
- Harms, A., Stanger, F. V., and Dehio, C. (2016b). Biological diversity and molecular plasticity of FIC domain proteins. *Annu. Rev. Microbiol.* 70, 341–360. doi: 10.1146/annurev-micro-102215-095245
- Harvey, S., and Greenwood, J. R. (1985). Isolation of *Campylobacter fetus* from a pet turtle. *J. Clin. Microbiol.* 21, 260–261.
- Hayes, F., and Van Melder, L. (2011). Toxins-antitoxins: diversity, evolution and function. *Crit. Rev. Biochem. Mol. Biol.* 46, 386–408. doi: 10.3109/10409238.2011.600437
- Jiang, M., Sullivan, S. M., Walker, A. K., Strahler, J. R., Andrews, P. C., and Maddock, J. R. (2007). Identification of novel *Escherichia coli* ribosome-associated proteins using isobaric tags and multidimensional protein identification techniques. *J. Bacteriol.* 189, 3434–3444.
- Karbstein, K. (2007). Role of GTPases in ribosome assembly. *Biopolymers* 87, 1–11. doi: 10.1002/bip.20762
- Kasari, V., Mets, T., Tenson, T., and Kaldalu, N. (2013). Transcriptional cross-activation between toxin-antitoxin systems of *Escherichia coli*. *BMC Microbiol.* 13:45. doi: 10.1186/1471-2180-13-45
- Kienesberger, S., Gorkiewicz, G., Joainig, M. M., Scheicher, S. R., Leitner, E., and Zechner, E. L. (2007). Development of experimental genetic tools for *Campylobacter fetus*. *Appl. Environ. Microbiol.* 73, 4619–4630. doi: 10.1128/AEM.02407-06
- Kienesberger, S., Sprenger, H., Wolfgruber, S., Halwachs, B., Thallinger, G. G., Perez-Perez, G. I., et al. (2014). Comparative genome analysis of *Campylobacter fetus* subspecies revealed horizontally acquired genetic elements important for virulence and niche specificity. *PLOS ONE* 9:e85491. doi: 10.1371/journal.pone.0085491
- Kinch, L. N., Yarbrough, M. L., Orth, K., and Grishin, N. V. (2009). Fido, a novel AMPylation domain common to fic, doc, and AvrB. *PLOS ONE* 4:e5818. doi: 10.1371/journal.pone.0005818
- Kirkbride, C. A., Gates, C. E., and Libal, M. C. (1989). Ovine and bovine abortion associated with *Fusobacterium nucleatum*. *J. Vet. Diagn. Invest.* 1, 272–273. doi: 10.1177/104063878900100318
- Lastovica, A. J., and Allos, B. M. (2008). “Clinical significance of *Campylobacter* and related species other than *Campylobacter jejuni* and *Campylobacter coli*,” in *Campylobacter*, 3 Edn, eds I. Nachamkin, C. M. Szymanski, and M. J. Blaser (Washington, DC: ASM), 123–149.
- Lehnher, H., Maguin, E., Jafri, S., and Yarmolinsky, M. B. (1993). Plasmid addition genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *J. Mol. Biol.* 233, 414–428. doi: 10.1006/jmbi.1993.1521
- Lepale, R., Geeraerts, D., Hallez, R., Guglielmini, J., Dreze, P., and Van Melder, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525. doi: 10.1093/nar/gkr131
- Liu, M., Zhang, Y., Inouye, M., and Woychik, N. A. (2008). Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5885–5890. doi: 10.1073/pnas.0711949105
- Lobato-Marquez, D., Diaz-Orejas, R., and Garcia-Del Portillo, F. (2016). Toxin-antitoxins and bacterial virulence. *FEMS Microbiol. Rev.* 40, 592–609. doi: 10.1093/femsre/fuw022
- Lobato-Marquez, D., Moreno-Cordoba, I., Figueroa, V., Diaz-Orejas, R., and Garcia-Del Portillo, F. (2015). Distinct type I and type II toxin-antitoxin modules control *Salmonella* lifestyle inside eukaryotic cells. *Sci. Rep.* 5:9374. doi: 10.1038/srep09374
- Logue, C. M., Sherwood, J. S., Elijah, L. M., Olah, P. A., and Dockter, M. R. (2003). The incidence of *Campylobacter* spp. on processed turkey from processing plants in the Midwestern United States. *J. Appl. Microbiol.* 95, 234–241. doi: 10.1046/j.1365-2672.2003.01969.x
- Maisonneuve, E., and Gerdes, K. (2014). Molecular mechanisms underlying bacterial persisters. *Cell* 157, 539–548. doi: 10.1016/j.cell.2014.02.050
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2009). Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* 4:19. doi: 10.1186/1745-6150-4-19
- Man, S. M. (2011). The clinical importance of emerging *Campylobacter* species. *Nat. Rev. Gastroenterol. Hepatol.* 8, 669–685. doi: 10.1038/nrgastro.2011.191
- Moreno, I., Codoner, F. M., Vilella, F., Valbuena, D., Martinez-Blanch, J. F., Jimenez-Almazan, J., et al. (2016). Evidence that the endometrial microbiota has an effect on implantation success or failure. *Am. J. Obstet. Gynecol.* 215, 684–703. doi: 10.1016/j.ajog.2016.09.075
- Mukherjee, S., Liu, X., Arasaki, K., McDonough, J., Galan, J. E., and Roy, C. R. (2011). Modulation of Rab GTPase function by a protein phosphocholine transferase. *Nature* 477, 103–106. doi: 10.1038/nature10335
- Norton, J. P., and Mulvey, M. A. (2012). Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *PLOS Pathog.* 8:e1002954. doi: 10.1371/journal.ppat.1002954
- Otter, A. (1996). *Fusobacterium necrophorum* abortion in a cow. *Vet. Rec.* 139, 318–319.
- Palanivelu, D. V., Goepfert, A., Meury, M., Guye, P., Dehio, C., and Schirmer, T. (2011). Fic domain-catalyzed adenylation: insight provided by the structural analysis of the type IV secretion system effector BepA. *Protein Sci.* 20, 492–499. doi: 10.1002/pro.581
- Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M. A., and Roy, C. R. (2008). Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320, 1651–1654. doi: 10.1126/science.1158160



- Park, S. J., Son, W. S., and Lee, B. J. (2013). Structural overview of toxin-antitoxin systems in infectious bacteria: a target for developing antimicrobial agents. *Biochim. Biophys. Acta* 1834, 1155–1167. doi: 10.1016/j.bbapap.2013.02.027
- Pascoe, B., Meric, G., Murray, S., Yahara, K., Mageiros, L., Bowen, R., et al. (2015). Enhanced biofilm formation and multi-host transmission evolve from divergent genetic backgrounds in *Campylobacter jejuni*. *Environ. Microbiol.* 17, 4779–4789. doi: 10.1111/1462-2920.13051
- Pelzer, E., Gomez-Arango, L. F., Barrett, H. L., and Nitert, M. D. (2017). Review: maternal health and the placental microbiome. *Placenta* 54, 30–37. doi: 10.1016/j.placenta.2016.12.003
- Rajashekara, G., Drozd, M., Gangaiah, D., Jeon, B., Liu, Z., and Zhang, Q. (2009). Functional characterization of the twin-arginine translocation system in *Campylobacter jejuni*. *Foodborne Pathog. Dis.* 6, 935–945. doi: 10.1089/fpd.2009.0298
- Ramage, H. R., Connolly, L. E., and Cox, J. S. (2009). Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLOS Genet.* 5:e1000767. doi: 10.1371/journal.pgen.1000767
- Ren, D., Kordis, A. A., Sonenshine, D. E., and Daines, D. A. (2014). The ToxAvapA toxin-antitoxin locus contributes to the survival of nontypeable *Haemophilus influenzae* during infection. *PLOS ONE* 9:e91523. doi: 10.1371/journal.pone.0091523
- Ren, D., Walker, A. N., and Daines, D. A. (2012). Toxin-antitoxin loci vapBC-1 and vapXD contribute to survival and virulence in nontypeable *Haemophilus influenzae*. *BMC Microbiol.* 12:263. doi: 10.1186/1471-2180-12-263
- Rollins, D. M., and Colwell, R. R. (1986). Viable but nonculturable stage of *Campylobacter jejuni* and its role in survival in the natural aquatic environment. *Appl. Environ. Microbiol.* 52, 531–538.
- Roy, C. R., and Cherfils, J. (2015). Structure and function of Fic proteins. *Nat. Rev. Microbiol.* 13, 631–640. doi: 10.1038/nrmicro3520
- Roy, C. R., and Mukherjee, S. (2009). Bacterial FIC proteins AMP up infection. *Sci. Signal.* 2:pe14. doi: 10.1126/scisignal.262pe14
- Saavedra De Bast, M., Mine, N., and Van Melder, L. (2008). Chromosomal toxin-antitoxin systems may act as antiaddiction modules. *J. Bacteriol.* 190, 4603–4609. doi: 10.1128/JB.00357-08
- Schrodinger, L. L. C. (2010). *The PyMOL Molecular Graphics System, Version 1.3r1*.
- Shao, Y., Harrison, E. M., Bi, D., Tai, C., He, X., Ou, H. Y., et al. (2011). TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Res.* 39, D606–D611. doi: 10.1093/nar/gkq908
- Shen, Z., Patil, R. D., Sahin, O., Wu, Z., Pu, X. Y., Dai, L., et al. (2016). Identification and functional analysis of two toxin-antitoxin systems in *Campylobacter jejuni*. *Mol. Microbiol.* 101, 909–923. doi: 10.1111/mmi.13431
- Skirrow, M. B., and Benjamin, J. (1980). '1001' *Campylobacters*: cultural characteristics of intestinal campylobacters from man and animals. *J. Hyg.* 85, 427–442.
- Smith, A. B., Lopez-Villarejo, J., Diago-Navarro, E., Mitchenall, L. A., Barendregt, A., Heck, A. J., et al. (2012). A common origin for the bacterial toxin-antitoxin systems *parD* and *ccd*, suggested by analyses of toxin/target and toxin/antitoxin interactions. *PLOS ONE* 7:e46499. doi: 10.1371/journal.pone.0046499
- Stanger, F. V., Burmann, B. M., Harms, A., Aragao, H., Mazur, A., Sharpe, T., et al. (2016). Intrinsic regulation of FIC-domain AMP-transferases by oligomerization and automodification. *Proc. Natl. Acad. Sci. U.S.A.* 113, E529–E537. doi: 10.1073/pnas.1516930113
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Wang, X., Lord, D. M., Hong, S. H., Peti, W., Benedik, M. J., Page, R., et al. (2013). Type II toxin/antitoxin MqsR/MqsA controls type V toxin/antitoxin GhoT/GhoS. *Environ. Microbiol.* 15, 1734–1744. doi: 10.1111/1462-2920.12063
- Wessner, F., Lacoux, C., Goeders, N., Fouquier d'Herouel, A., Matos, R., Serron, P., et al. (2015). Regulatory crosstalk between type I and type II toxin-antitoxin systems in the human pathogen *Enterococcus faecalis*. *RNA Biol.* 12, 1099–1108. doi: 10.1080/15476286.2015.1084465
- Woolery, A. R., Luong, P., Broberg, C. A., and Orth, K. (2010). AMPylation: something old is new again. *Front. Microbiol.* 1:113. doi: 10.3389/fmicb.2010.00113
- Worby, C. A., Mattoo, S., Kruger, R. P., Corbeil, L. B., Koller, A., Mendez, J. C., et al. (2009). The fic domain: regulation of cell signaling by adenylation. *Mol. Cell.* 34, 93–103. doi: 10.1016/j.molcel.2009.03.008
- Xiao, J., Worby, C. A., Mattoo, S., Sankaran, B., and Dixon, J. E. (2010). Structural basis of Fic-mediated adenylation. *Nat. Struct. Mol. Biol.* 17, 1004–1010. doi: 10.1038/nsmb.1867
- Yamaguchi, Y., and Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat. Rev. Microbiol.* 9, 779–790. doi: 10.1038/nrmicro2651
- Yang, M., Gao, C., Wang, Y., Zhang, H., and He, Z. G. (2010). Characterization of the interaction and cross-regulation of three *Mycobacterium tuberculosis* RelBE modules. *PLOS ONE* 5:e10672. doi: 10.1371/journal.pone.0010672
- Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., and Orth, K. (2009). AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* 323, 269–272. doi: 10.1126/science.1166382
- Zechner, E. L., Lang, S., and Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1073–1087. doi: 10.1098/rstb.2011.0207
- Zekarias, B., Mattoo, S., Worby, C., Lehmann, J., Rosenbusch, R. F., and Corbeil, L. B. (2010). *Histophilus somni* IbpA DR2/Fic in virulence and immunoprotection at the natural host alveolar epithelial barrier. *Infect. Immun.* 78, 1850–1858. doi: 10.1128/IAI.01277-09
- Zhang, Y., and Inouye, M. (2011). RatA (YfjG), an *Escherichia coli* toxin, inhibits 70S ribosome association to block translation initiation. *Mol. Microbiol.* 79, 1418–1429. doi: 10.1111/j.1365-2958.2010.07506.x
- Zhu, L., Sharp, J. D., Kobayashi, H., Woychik, N. A., and Inouye, M. (2010). Noncognate *Mycobacterium tuberculosis* toxin-antitoxins can physically and functionally interact. *J. Biol. Chem.* 285, 39732–39738. doi: 10.1074/jbc.M110.163105

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RO and handling Editor declared their shared affiliation.

Copyright © 2017 Sprenger, Kienesberger, Pertschy, Pörtl, Konrad, Bhutada, Vorkapic, Atzmüller, Feist, Högenauer, Gorkiewicz and Zechner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Disulfide Bond in the Membrane Protein IgaA Is Essential for Repression of the RcsCDB System

M. Graciela Pucciarelli<sup>1,2,3†</sup>, Leticia Rodríguez<sup>1†</sup> and Francisco García-del Portillo<sup>1\*</sup>

<sup>1</sup> Laboratorio de Patógenos Bacterianos Intracelulares, Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas (CNB-CSIC), Madrid, Spain, <sup>2</sup> Departamento de Biología Molecular, Universidad Autónoma de Madrid, Madrid, Spain, <sup>3</sup> Centro de Biología Molecular Severo Ochoa-Consejo Superior de Investigaciones Científicas (CBMSO-CSIC), Madrid, Spain

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Sultan Zainal Abidin University,  
Malaysia

### Reviewed by:

Nadim Majdalani,  
National Institutes of Health (NIH),  
United States  
Kevin D. Young,  
University of Arkansas for Medical  
Sciences, United States

### \*Correspondence:

Francisco García-del Portillo  
fgportillo@cnb.csic.es

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 September 2017

**Accepted:** 14 December 2017

**Published:** 22 December 2017

### Citation:

Pucciarelli MG, Rodríguez L and  
García-del Portillo F (2017) A  
Disulfide Bond in the Membrane  
Protein IgaA Is Essential  
for Repression of the RcsCDB  
System. *Front. Microbiol.* 8:2605.  
doi: 10.3389/fmicb.2017.02605

IgaA is an integral inner membrane protein that was discovered as repressor of the RcsCDB phosphorelay system in the intracellular pathogen *Salmonella enterica* serovar Typhimurium. The RcsCDB system, conserved in many members of the family *Enterobacteriaceae*, regulates expression of varied processes including motility, biofilm formation, virulence and response to envelope stress. IgaA is an essential protein to which, in response to envelope perturbation, the outer membrane lipoprotein RcsF has been proposed to bind in order to activate the RcsCDB phosphorelay. Envelope stress has also been reported to be sensed by a surface exposed domain of RcsF. These observations support a tight control of the RcsCDB system by RcsF and IgaA via mechanisms that, however, remain unknown. Interestingly, RcsF and IgaA have four conserved cysteine residues in loops exposed to the periplasmic space. Two non-consecutive disulfide bonds were shown to be required for RcsF function. Here, we report mutagenesis studies supporting the presence of one disulfide bond (C404-C425) in the major periplasmic loop of IgaA that is essential for repression of the RcsCDB phosphorelay. Our data therefore suggest that the redox state of the periplasm may be critical for the control of the RcsCDB system by its two upstream regulators, RcsF and IgaA.

**Keywords:** *Salmonella*, IgaA, periplasmic domain, cysteine, disulfide bond, RcsCDB

## INTRODUCTION

The RcsCDB phosphorelay is a regulatory system conserved in most members of the family *Enterobacteriaceae* (Majdalani and Gottesman, 2007). A major role of this system is to monitor cell envelope stress, responding to alterations in outer membrane integrity and peptidoglycan structure (Farris et al., 2010; Evans et al., 2013; Konovalova et al., 2016). The tripartite RcsCDB system is atypical compared to most known phosphorelays, normally progressing from a sensor membrane protein to a cytosolic response regulator (Wolanin et al., 2002; Majdalani and Gottesman, 2005). In the RcsCDB system, the signal is transmitted from the sensor inner membrane protein RcsC to the intermediate membrane protein RcsD to end with phosphorylation of a conserved aspartate residue in the RcsB response regulator. The RcsCDB system controls expression of more than 40 genes involved in biofilm formation, synthesis of exopolysaccharide capsule, motility, and virulence among others (Hagiwara et al., 2003; Majdalani and Gottesman, 2007;

Mariscotti and Garcia-del Portillo, 2009; Howery et al., 2016). The genes of the RcsCDB regulon were initially classified in those regulated exclusively by RcsB and, a second group including those involved in exopolysaccharide synthesis, which are controlled by RcsB and the co-regulator RcsA (Dierksen and Trempey, 1996; Navasa et al., 2013). Recent studies in *Escherichia coli* demonstrate that RcsB can heterodimerize with other co-regulatory proteins (Pannen et al., 2016). RcsB has also been shown to have a notable conformational dynamism (Casino et al., 2017), which could explain its capacity for providing different responses depending the phosphorylation status and the type and intensity of the stimulus (Mariscotti and Garcia-del Portillo, 2009; Latasa et al., 2012).

The RcsCDB system displays a feature conserved in most other regulatory networks, regarding its rapid response to the stress signal followed by a progressive decrease in activity once the bacterium adapts to the new environmental conditions (Gao and Stock, 2017). Of interest, these regulatory systems are “prepared to act” as denoted by the presence of all components of the signaling cascade even in the absence of stimulus. Thus, isogenic mutants of *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) displaying differences in the expression of RcsB target genes produce similar relative levels of the RcsC, RcsD, and RcsB proteins (Dominguez-Bernal et al., 2004).

Two important regulatory elements acting upstream of the RcsCDB system are the outer membrane lipoprotein RcsF and the integral inner membrane protein IgaA. RcsF was first reported as a lipoprotein that transmits a stress signal to the inner membrane sensor RcsC following cell envelope perturbations (Majdalani et al., 2005). IgaA was discovered as an integral inner membrane protein that contributes to attenuate the growth rate of *S. Typhimurium* inside eukaryotic cells (Cano et al., 2001). Subsequent studies revealed that the mucoid phenotype exhibited by a mutant bearing a R188H mutation in IgaA was linked to over-activation of the RcsCDB phosphorelay (Cano et al., 2002; Dominguez-Bernal et al., 2004). IgaA is predicted to have four transmembrane domains with the R188 residue located in one of the cytosolic domains (Dominguez-Bernal et al., 2004). Unlike the wild type IgaA protein, produced at constant levels in actively growing and resting bacteria, the R188H variant is unstable in stationary phase (Dominguez-Bernal et al., 2004). Although the loss of IgaA can be supported in non-growing bacteria, genetic evidence obtained in *S. Typhimurium* and *E. coli* demonstrates that *igaA* is an essential gene (Cano et al., 2002; Cho et al., 2014). Of note, IgaA become dispensable if the RcsCDB system is genetically inactivated (Cano et al., 2002). Moreover, loss-of-function mutations in the RcsCDB system are selected at high rate when attempting to replace the wild-type *igaA* gene by a null allele (Mariscotti and Garcia-Del Portillo, 2008). Altogether, these observations reveal a critical function of IgaA as a dedicated repressor of the RcsCDB phosphorelay in actively growing bacteria. Transcriptomic analyses also pointed to a major role of IgaA in fine-tuning the RcsCDB phosphorelay (Mariscotti and Garcia-del Portillo, 2009).

A recent study has provided the first insights into the mechanism by which the upstream regulators, RcsF and IgaA, could control activity of the RcsCDB phosphorelay (Cho

et al., 2014). These authors showed that in steady-state growth conditions, RcsF is exposed in the external face of the outer membrane via interaction with OmpA and BamA, the major component of the  $\beta$ -barrel assembly machinery. Following peptidoglycan stress, RcsF fails to interact with OmpA/BamA and, as a result, retained in the periplasmic space. In this condition RcsF binds to the major periplasmic domain of IgaA to activate the RcsCDB phosphorelay (Cho et al., 2014). Based on the previous functional data obtained with IgaA, the RcsF-IgaA interaction must therefore alleviate the repression that IgaA exerts on the RcsCDB system in non-stimulatory conditions. Envelope stress can also be directed sensed by the surface-exposed domain of RcsF when defects in lipopolysaccharide structure occur (Konovalova et al., 2016).

RcsF has four conserved cysteines that form disulfide bonds (Leverrier et al., 2011). The formation of these disulfide bonds in RcsF depends on DsbC, the main disulfide isomerase, which together with the disulfide oxidase DsbA, control the formation and correct configuration of disulfide bonds (Denoncin and Collet, 2013). Disulfide bridges can play a structural role, as stable bonds, or; alternatively, contribute to catalysis by forming reversible disulfide bonds in the catalytic site (Denoncin et al., 2013). This latter case is exemplified by periplasmic oxidoreductases such as DsbA and DsbC. Whether the disulfide bonds of RcsF play a role in the interaction with IgaA is unknown.

In this study, we investigated the presence of disulfide bonds in IgaA, which contains four conserved cysteine residues in its major periplasmic domain. Our results are consistent with the presence of a disulfide bond in IgaA that is important for its function as repressor of the RcsCDB phosphorelay.

## MATERIALS AND METHODS

### Bacterial Strains and Growth Conditions

The bacterial strains and plasmid used in this study are listed in Supplementary Table S1. Bacteria were cultured in Luria-Bertani (LB) broth at 37°C in shaking (150 rpm) conditions. To prepare material from mid-exponential cultures, the overnight culture was diluted 1:100 in fresh LB medium and collected at optical density (absorbance at 600 nm) of  $\sim 0.2$ – $0.3$ . The remaining culture was incubated for additional 18 h to obtain stationary phase cultures. When required, the medium was supplemented with ampicillin (50  $\mu$ g/ml), kanamycin (30  $\mu$ g/ml), tetracycline (10  $\mu$ g/ml), or chloramphenicol (10  $\mu$ g/ml).

### Mutagenesis of Periplasmic Cysteines

The mutagenesis was carried out with the Quick-change™ site-directed mutagenesis kit from Stratagene, following manufacture recommendations. The oligonucleotides used for these procedures, including those degenerated introducing the desired point mutations, are listed in Supplementary Table S2. A copy of the *S. Typhimurium* wild-type *igaA* gene was cloned in the pBAD18 vector (plasmid pNG1062, Supplementary Table S1) and used as template for the mutagenesis kit to obtain pNG1062-derivate plasmids containing the mutant alleles (C404S, C425S, C498, and C504S), which were cloned in

*E. coli* DH5 $\alpha$  (Supplementary Table S1). The desired mutations were confirmed by sequencing. To generate the C404S–498S and C404S–C504S mutant alleles, a BspI/BspI fragment from the pLR1435 [pBAD18::*igaA*(C404S)] was used to replace the same region in pLR1438 [pBAD18::*igaA*(C498S)] and pLR1481 [pBAD18::*igaA*(C504S)] plasmids. The series of pNG1062-derivate plasmid (pBAD18 backbone) was transferred to the *S. Typhimurium* MD0835 strain [*igaA2*::KXX  $\Delta$ (*apbE'*-*rcsC'*)], a mutant not producing IgaA and with an additional mutation in *rcsC* (Mariscotti and Garcia-Del Portillo, 2008). The production of the distinct IgaA variants with mutated cysteine residues was confirmed in the MD0835-derivate strains grown in LB-0.2% L-arabinose and subsequent analysis of total protein extracts by Western assay using anti-IgaA antibody (Cano et al., 2002).

## Generation of *S. Typhimurium* Strains with *igaA* Mutant Alleles (C404S, C425S, C498S, C504S) Disposed in the Chromosome

To transfer the *igaA* mutant alleles to the chromosome, they were first moved from the pBAD cloning vector to the pCVD442 suicide vector (Donnenberg and Kaper, 1991), which has the counter selectable marker *sacB*, using *E. coli* DH5 $\alpha$  as host strain (Supplementary Table S1). Since we did not initially know whether the *igaA* mutant alleles could support viability (*igaA* is an essential gene in a *rcsCDB*<sup>+</sup> background), the *igaA* mutant alleles were first moved to the chromosome of *S. Typhimurium* strain MD1446 (*igaA2*::KXX *zhf-6311*::Tn10dTet *rcsC*::MudQ). This was done by conjugation using as donor *E. coli* SM10 $\lambda$ pir carrying the respective series of pCVD442 derivate plasmids. Loss of kanamycin resistance in the *S. Typhimurium* recipient strain was indicative of replacement of the *igaA* null allele (*igaA2*::KXX). The cysteine-defective *igaA* alleles were further transferred by P22 phage transduction to a clean wild-type background selecting by tetracycline resistance (flanking marker *zhf-6311*::Tn10dTet). The *igaA* gene was PCR-amplified from all Tet<sup>R</sup> transductants to confirm presence in the chromosome of the mutation in the codon of the corresponding cysteine residue. Whereas all single *igaA* mutants (C404S, C425S, C498S, and C504S) proved to support viability, no transductants were

obtained attempting to transduce to a wild-type background the double mutant alleles C404S–C498S and C404S–C504S. This result was consistent with the viability test performed with the different *igaA* variants expressed from pBAD using different L-arabinose concentrations (see below, **Table 1**).

## Western Blot Analyses

Preparation of protein extracts, electrophoresis and Western assay conditions using polyclonal rabbit anti-IgaA antibody were as described (Dominguez-Bernal et al., 2004).

## AMS Alkylation Assays

These assays were performed using the 4'-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMS) reagent, as described (Jurado et al., 2006).

## $\beta$ -galactosidase Assays

Levels of  $\beta$ -galactosidase derived from the *gmm::lacZ* transcriptional fusion were assayed as described by Miller, following the chloroform/SDS permeabilization procedure (Miller, 1972). For these assays, bacteria were grown in LB medium to mid-exponential phase (OD<sub>600</sub> ~ 0.2–0.3).

## Motility Assays

Motility of the different strains used was monitored by motility assays in soft agar plates, as described (Rosu et al., 2006).

## Statistical Analysis

Data were analyzed by one-way ANOVA using Prism version 5.0 (Graph-Pad Software). Differences in values with *P* < 0.05 were considered significant.

# RESULTS

## IgaA Has Four Cysteines in the Periplasmic Domain Conserved in All Orthologs of Enteric Bacteria

Our previous studies in *S. Typhimurium* showed that IgaA is an inner membrane protein produced at relatively constant

**TABLE 1 |** Suppression of lethality associated to the *igaA*::Km null allele by ectopic expression of different IgaA variants.

% L-arabinose (w/v)	Number of Km <sup>R</sup> transductants obtained with an <i>igaA</i> ::Km P22 phage lysate							
	pBAD18 (empty vector)	<i>igaA</i> <sup>+</sup> pBAD18	<i>igaA</i> -C404S	<i>igaA</i> -C425S	<i>igaA</i> -C498S	<i>igaA</i> -C504S pBAD18	<i>igaA</i> -C404S-C498S pBAD18	<i>igaA</i> -C404S-C504S pBAD18
0	0	0	0	0	0	0	0	0
0.002	0	17/24	0	0	0	0	0	0
0.02	0	1250/1440	389/369*	437/384*	728/726*	384/390*	0	0
0.2	0	5500/6000	505/454*	393/360	716/720	408/412	484/500*	318/286*
2	0	10500/9000	687/612*	716/717	970/1138	600/650	543/672*	406/454*
5	0	3200/3400	850	618/716	768/815	730	540/672*	492/315*

\*These transductants were mucoid, denoting reduced repression of the *RcsCDB* system by the respective IgaA variant. No Km<sup>R</sup> transductants bearing the *igaA*::Km null allele were obtained when using 2% glucose (w/v) instead of arabinose.



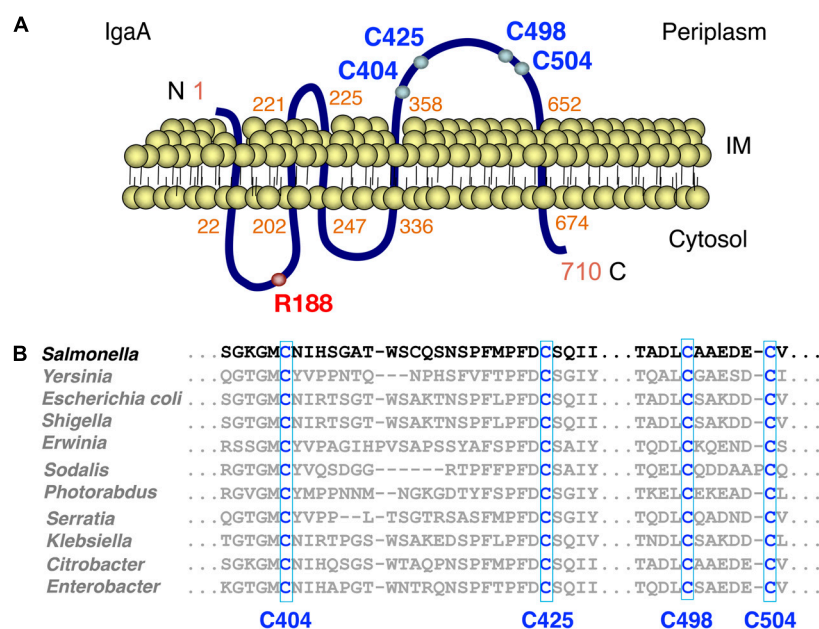
levels in all growing conditions (Dominguez-Bernal et al., 2004). Programs that predict transmembrane helix regions and protein topology (THMM, SOSUI, TMPred, PredictProtein) indicate that the 710 amino acid protein IgaA has five transmembrane segments, resulting in two cytosolic domains (residues 22–202, 247–336), one small periplasmic loop (221–225) and one major periplasmic domain (residues 358–652) (Figure 1A). Mutations in specific residues of the two cytosolic loops (R188H, T191P, G262R) as well as in the periplasmic domain (L514P, L643P) impact negatively the capacity of IgaA to repress the RcsCDB phosphorelay (Dominguez-Bernal et al., 2004). We further noted that the periplasmic domain of *S. Typhimurium* IgaA has four cysteines, (C404, C425, C498, C504) highly conserved in IgaA orthologs found in distinct genera of enteric bacteria (Figure 1B). Based on this observation, we assessed whether these periplasmic cysteines could form disulfide bonds and play an important role in function.

To determine the role played by the C404, C425, C498, and C504 residues, we generated isogenic *S. Typhimurium* mutants lacking each of these cysteines. To this aim, we first introduced in the chromosome the corresponding *igaA* point mutant allele using as recipient an *igaA::KXX rcsC* strain, to subsequently pass the allele to a wild-type (*rcsCDB*<sup>+</sup>) genetic background by P22 phage transduction (see section “Material and Methods”). Importantly, none of the C404S, C425C, C498S, and C504S mutations affected protein stability in actively growing bacteria. Thus, in bacteria grown to exponential phase these IgaA variants were detected with similar levels than those of wild-type IgaA or the previously characterized R188H variant

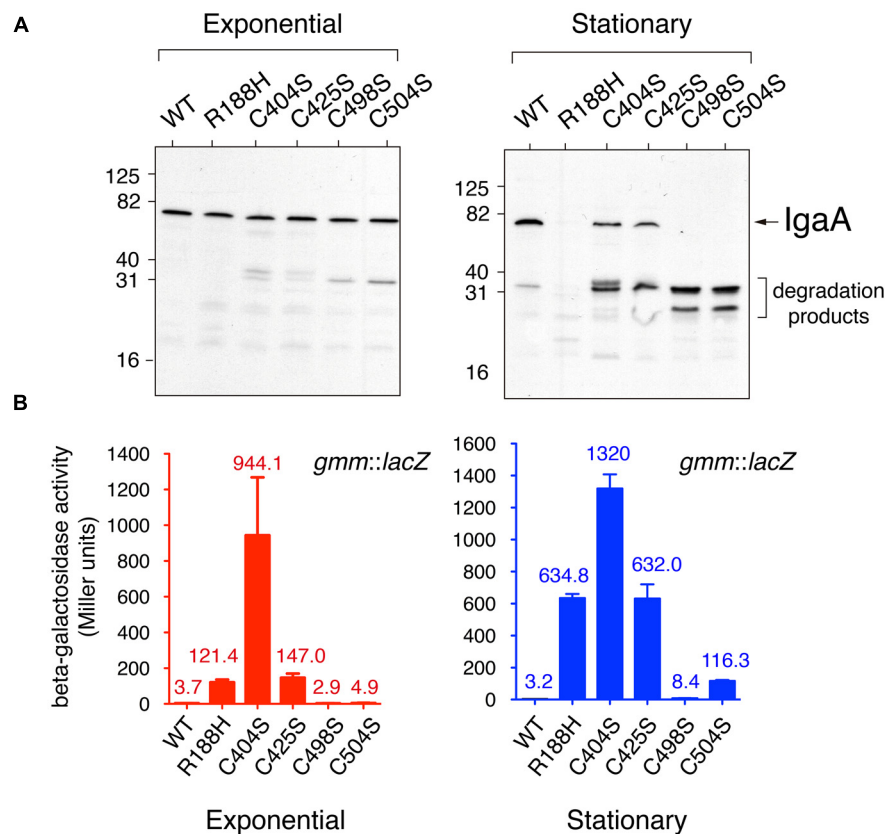
(Figure 2A) (Dominguez-Bernal et al., 2004). Nonetheless, all these four cysteine variants (C404S, C425S, C498S, and C504S) were unstable in stationary phase (Figure 2A). This phenomenon was reminiscent of that observed for other partially inactive variants such as R188H and L514P (Dominguez-Bernal et al., 2004). Therefore, we concluded that the elimination of any of the four conserved residues (C404, C425, C498, and C504) may lead to structural changes in IgaA that affect its stability when bacteria reach stationary phase. Considering the essentiality of IgaA linked to necessary repression of the RcsCDB system, all these cysteine variants were, however, expected to retain some partial function in growing bacteria.

## Elimination of the Periplasmic Cysteines (C404, C425, C498, and C504) of IgaA Results in Distinct De-repression Levels of the RcsCDB System

To determine the capacity of the IgaA cysteine variants to repress the RcsCDB phosphorelay, we monitored several phenotypic traits associated to the activity of this regulatory system. We included production of colanic acid capsule, which is positively controlled by RcsCDB; and, flagella production, negatively regulated by the system. In a first series of experiments, we measured in actively growing (exponential) and resting (stationary phase) bacteria the expression levels of *gmm* (*wcaH*), a gene encoding GDP-mannose mannosyl hydrolase, an enzyme involved in colanic acid capsule synthesis. The results showed that the C404S



**FIGURE 1 |** The integral inner membrane protein IgaA has four conserved periplasmic cysteines. **(A)** Predicted topology of the *Salmonella Typhimurium* IgaA protein of 710 amino acids length. The periplasmic cysteines C404, C425, C498, and C504 as well as the cytosolic R188 residue important for IgaA function, are shown. **(B)** Alignment of the periplasmic domain of distinct IgaA orthologs from the indicated members of the *Enterobacteriaceae* family. Alignment was performed with the Clustal Omega tool.

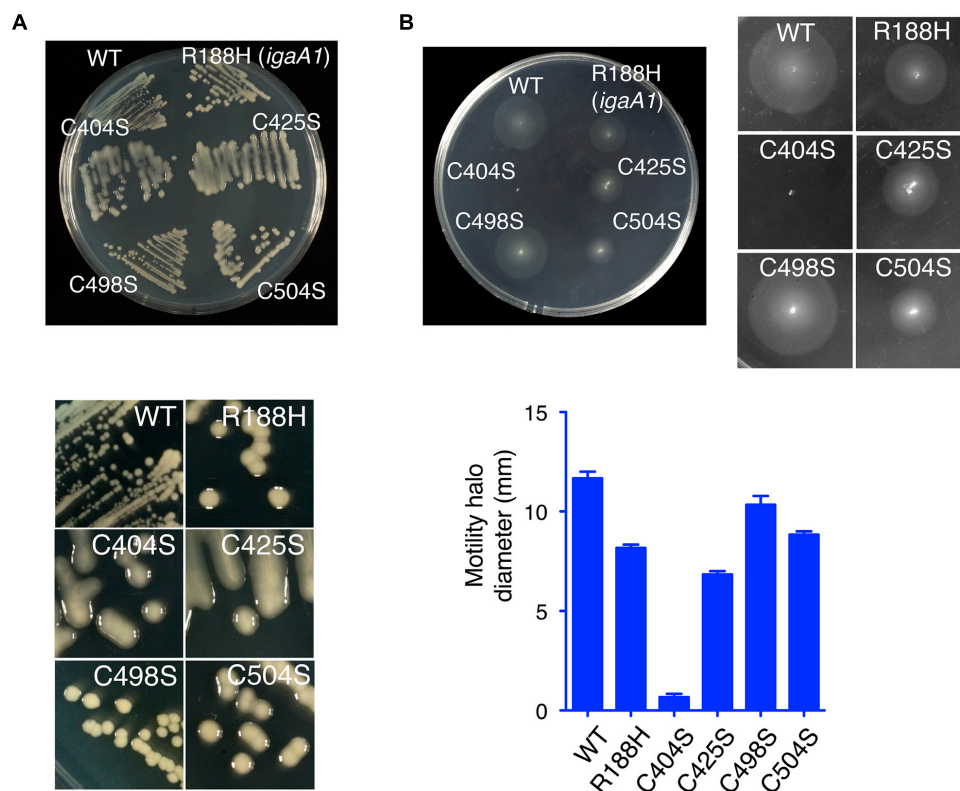


**FIGURE 2 |** *Salmonella* Typhimurium IgaA variants in conserved cysteines show distinct capacities to repress the RcsCDB system. **(A)** IgaA was detected in western assays using samples obtained from bacteria grown in LB medium at 37°C to mid-exponential (OD<sub>600</sub> = 0.2) or stationary (OD<sub>600</sub> = 2.0) phase. Note the degradation of the IgaA cysteine variants when bacteria reach the stationary phase. The previously characterized IgaA variant R188H (Dominguez-Bernal et al., 2004) was used as control of protein unstable in stationary phase. Numbers on the left indicate the position of molecular weight markers (in kDa). **(B)** Activity of the RcsCDB system monitored with a *gmm::lacZ* reporter fusion in bacteria growing to exponential or stationary phases. *gmm* encodes an enzyme involved in synthesis of the colanic acid capsule and is positively regulated by the RcsCDB system. Data are the average and standard deviation of three independent experiments. The averages are shown as numbers on top of the respective bars.

mutation caused partial de-repression of the RcsCDB system, even in actively growing bacteria when the protein remained stable (Figures 2A,B). Among the other mutants, a gradual variation in the level of RcsCDB activity was noted following the order C404S > C425S ~ R188H > C504S > C498S > wild-type (Figure 2B). Interestingly, the levels of the RcsCDB system in stationary phase inferred from the *gmm::lacZ* reporter fusion increased only slightly despite the partial degradation observed for some of the IgaA mutant proteins such as C404S, C425S, C498S, or C504S (Figures 2A,B). These data are consistent with a major role of IgaA in repressing the RcsCDB that is critical only during active growth. These data also indicated that C404 and C425 are cysteines more important for function in comparison to C498 and C504.

Additional phenotypic traits that were examined included the formation of mucoid colonies on plates (signal of capsule formation) and motility assays in soft agar plates. In agreement with the data obtained with the *gmm(wcaH)::lacZ* reporter fusion, the mutant producing IgaA-C404S was highly mucoid and non-motile (Figures 3A,B). For the rest of mutants, a gradual

variation in RcsCDB activity was noted in the mucoidity and motility tests (Figures 3A,B), which in some cases were not completely matching the *gmm::lacZ* assays performed in liquid culture (Figure 2B). Thus, although the IgaA-C504S variant exhibited lower *gmm::lacZ* expression than R188H (Figure 2B), bacteria producing this C504S variant were slightly more mucoid on plates (Figure 3A). R188H, C425S, and C504S variants also displayed an intermediate phenotype in motility despite their variations in the *gmm::lacZ* assays or mucoidity on plates (Figures 2B, 3A,B). Such discrepancies in the different tests may be influenced by the different growth conditions used -liquid culture vs. solid agar media plates-. Interestingly, mutations in defined RcsB residues alter the phosphorylation status of this regulator with consequences in either mucoidity or motility, but not in both phenotypic traits (Casino et al., 2017). Some of the mutations described here in the periplasmic cysteines of IgaA could result in distinct RcsB~P/RcsB ratios, a hypothesis to be tested in future studies. Despite the minor phenotypic differences noted in IgaA for the R188H, C425S, and C504S variants; taken together, our data support that among the four periplasmic



**FIGURE 3 |** The periplasmic cysteines C404 and C425 are essential for the negative regulation that IgaA imposes over the RcsCDB phosphorelay. **(A)** Mucoid phenotype of strains producing the C404S, C425S, C504S, and R188H variants. **(B)** Motility test in soft agar. Note that the high repression of motility in the strain producing the C404S variant. A graphic showing motility halo diameter (average and standard deviation from three independent assays) is also depicted.

cysteines analyzed, C404 and C425 are residues with a more critical role for repression of the RcsCDB phosphorelay.

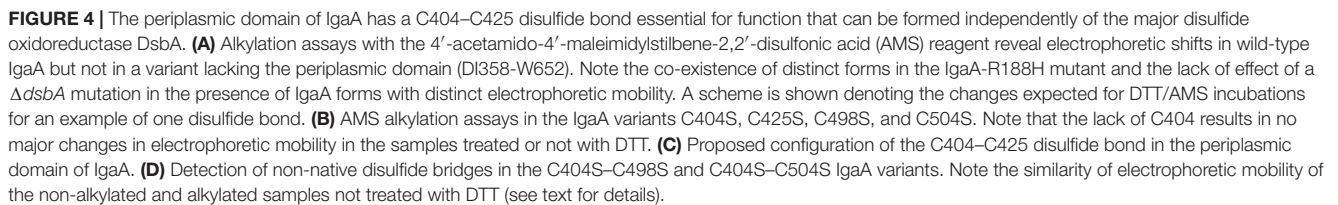
## IgaA Has One Disulfide Bond in the Periplasmic Domain

To elucidate whether the contribution of the periplasmic cysteines to IgaA function relies in the formation of disulfide bonds, we carried out alkylation experiments using the 4'-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMS) reagent (Jurado et al., 2006; Denoncin et al., 2013). These experiments were performed in exponential phase ( $OD_{600} \sim 0.2$ – $0.3$ ), in which all IgaA variants with mutated cysteines are stable (see **Figure 2A**). AMS is a maleimide compound that binds to free thiol groups. Thus, it is possible to differentiate the presence of a disulfide bond if changes in electrophoretic mobility are detected in samples incubated with AMS and previously treated or not with a reducing agent such as di-thio-threitol (DTT). When *S. Typhimurium* wild type cells were incubated in solutions with or without DTT and further treated or not with AMS, we detected four IgaA forms with distinct electrophoretic mobility (**Figure 4A**). This result proved the presence of at least one disulfide bond in IgaA.

Besides the four periplasmic cysteines, IgaA of *S. Typhimurium* has four additional cysteines that map in

the first and second transmembrane regions (C14, C219, respectively), the second cytosolic domain (C259) and, the short cytosolic domain encompassing the C-terminal end of the protein (C697). To discard the contribution of these cysteines to the mobility shift displayed by wild-type IgaA in the alkylation assays, we generated a variant lacking the periplasmic domain ( $\Delta I358$ -W652). This variant, unlike the full-length protein, did not exhibit electrophoretic shift in cells exposed to DTT and subsequently to AMS (**Figure 4A**). Therefore, none of the non-periplasmic cysteines of IgaA contribute to the formation of disulfide bonds. Interestingly, the different electrophoretic forms of IgaA observed in the AMS alkylation assays in wild-type bacteria were also detected in a  $\Delta dsbA$  mutant (**Figure 4A**). This result indicated that the disulfide bond present in IgaA can be formed in the absence of the major disulfide oxidase DsbA.

To define the configuration of the periplasmic disulfide bond(s) inferred in wild-type IgaA, we next performed AMS alkylation assays in isogenic strains lacking each of the four conserved periplasmic cysteines. These assays revealed an important contribution of C404 in the formation of a disulfide bond. Thus, unlike the C425S, C498S, and C504S variants, the C404S variant did not migrate differently in the presence/absence of DTT (**Figure 4B**). Interestingly, the four variants in the conserved periplasmic cysteines (C404, C425,



Taken together, these data are consistent with a model in which the function of IgaA as repressor of the RcsCDB system depends on the redox state of its periplasmic domain, which in its functional conformation may involve the formation of a C404–C425 disulfide bond (**Figure 4C**).



## The Lack of the C404–C425 Disulfide Bond or the Formation of Alternative Non-native Disulfide Bonds Are Lethal in a RcsCDB<sup>+</sup> Background

To further support the essential role that the C404–C425 disulfide bond plays in repression of the RcsCDB phosphorelay, we generated IgaA double mutants. Our aim was to affect the native bridge (C404–C425) and to simultaneously impair alternative non-native disulfide bonds predicted by the AMS alkylation assays and involving either C498 or C504 (**Figure 4B**). Therefore, the new IgaA variants were C404S–C498S and C404S–C504S. The AMS alkylation assays revealed no difference in migration between the non-alkylated and alkylated samples not exposed to DTT for any of these two double mutants (**Figure 4D**). This result, much more evident in the case of the C404S–C498S variant, supported the absence of free periplasmic cysteines and, therefore, the presence a “non-native” disulfide bond (C425–C504). In the C404S–C504S variant, the alkylation assays revealed two bands in the case of the samples non-treated with DTT and exposed to AMS (**Figure 4D**), implying the co-existence of molecules with and without a non-native C425–C498 disulfide bond. This phenomenon resembled at some extent the co-existence of molecules with different redox state observed for the IgaA-R188H variant (**Figure 4A**).

We next assessed whether the non-native C425–C498 or C425–C504 bonds could provide functionality to the protein. To this aim, we analyzed the capacity of the different IgaA variants to suppress lethality associated to the presence of an *igaA::Km* null allele in an RcsCDB<sup>+</sup> genetic background (Cano et al., 2002; Mariscotti and Garcia-Del Portillo, 2008). The number of *igaA::Km* transductants obtained in strains bearing inducible pBAD18 vectors expressing the different IgaA variants was determined in the absence/presence of the inducer, L-arabinose. As a negative control, we used a strain with the pBAD18 empty vector, for which no Km<sup>R</sup> transductants carrying the *igaA::km* null allele were obtained regardless the absence/presence of inducer.

These assays showed that the C404S–C498S and C404S–C504S variants were not capable of suppressing the lethality associated to the *igaA::km* mutant when induced at 0.02% L-arabinose, a concentration sufficient to prevent lethality by any of the single mutants lacking each of the conserved cysteine residues (**Table 1**). Furthermore, we observed that when suppressing lethality at higher arabinose concentrations, all transductants were mucoid (**Table 1**). This result was indicative of the extremely limited function of these double C404S–C498S and C404S–C504S variants, even when produced at high levels. Therefore, the non-native C425–C498 and C425–C504 disulfide bonds are not optimal to provide function to IgaA.

## DISCUSSION

In this study, we have examined the role in function of four conserved cysteines located in the periplasmic domain of the

RcsCDB repressor IgaA. Our data prove the presence in the native protein of a periplasmic disulfide bond in the configuration C404–C425. The alkylation experiments suggest that a C404S mutation renders the protein unable to form any stable disulfide bond, a scenario slightly different to that of the C425S mutation, in which non-native C404–C498 or C404–C504 bonds were inferred. This interpretation agrees with the high de-repression of the RcsCDB phosphorelay observed in *S. Typhimurium* strains expressing the IgaA–C404S variant. Importantly, our data discarded any compensatory role in function for the non-native disulfide bonds C404–C498 or C404–C504 that apparently occur in the C425S mutant. Therefore, not all disulfide bonds capable of forming in the periplasmic domain support equally IgaA function as RcsCDB repressor. A similar conclusion was reached for RcsF of *E. coli*, which has two not functionally equivalent disulfide bonds (Leverrier et al., 2011; Rogov et al., 2011), with one of them proposed to be more relevant for function (Leverrier et al., 2011).

The alkylation experiments performed with the  $\Delta dsbA$  mutant discarded an absolute requirement of this disulfide oxidase for formation of the C404–C425 bond. Thus, different IgaA forms with distinct electrophoretic mobility were detected in this  $\Delta dsbA$  mutant depending the presence/absence of DTT and/or AMS. This result opens the possibility of IgaA being recognized by alternative disulfide oxidases. Besides the pair DsbA/DsbB, *S. Typhimurium* encodes the paralogs DsbL and DsbI (Lin et al., 2009) and has an additional DsbA paralog, SrgA, encoded in the virulence plasmid (Bouwman et al., 2003). Interestingly, SsaC, a virulence-related protein related to the type III secretion system encoded in the *Salmonella* pathogenicity island 2 (SPI-2), was shown to be oxidized indistinctly by DsbA or SrgA (Miki et al., 2004). Further experiments are needed to confirm whether IgaA could be recognized by any of these alternative disulfide oxidases.

An unexpected finding of our study was the co-existence of IgaA molecules with different redox states. This situation, not observed for the wild-type IgaA, was evident for the R188H and the C404S–C504S variants (**Figures 4A,D**). Such behavior may reflect conformational plasticity in the periplasmic domain of IgaA, which may be altered at some degree when mutations in key residues important for function are introduced. Importantly, the observations with the R188H mutation proved that disrupting a key residue for function in the cytosolic side can have consequences in the structure of the periplasmic loop. This evidence points to signal transmission within the IgaA molecule from one to other side of the inner membrane.

Another aspect of interest found in the study was the instability of all IgaA variants lacking the conserved periplasmic cysteines when bacteria reached stationary phase. Proteolysis in stationary phase is not a common process. Some of the few cases known include YfgM and the formate dehydrogenase subunit FdoH, two substrates of the cytosolic FtsH protease (Westphal et al., 2012; Bittner et al., 2015). Based on our findings with the mutated IgaA variants, it is tempting to speculate on a protease that could monitor IgaA for proper folding in stationary phase.

Noteworthy, mass spectrometry analyses revealed RcsF and the protease DegP as partners interacting with the periplasmic domain of IgaA, (Cho et al., 2014). Whether DegP monitors correct folding of the IgaA periplasmic domain requires further experiments.

An important feature of all cysteine IgaA variants is that they were fully stable in actively growing bacteria (Figure 2), a condition in which de-repression of the RcsCDB system (measured by the *gmm::lacZ* fusion) was evident in some cases as those of C404S and C425S mutations (Figure 2B). This experimental evidence supports a mode of action of IgaA involving a C404–C425 disulfide bond, absolutely essential for repressing the RcsCDB system.

The gradation observed in the phenotypic assays involving mucoidity and motility and the discrepancy found for some of the IgaA variants tested (R188H, C425S, C504S), may reflect a functional link between IgaA and the conformational plasticity recently reported for RcsB (Casino et al., 2017). RcsB plasticity, which is directly linked to its phosphorylation status, could allow the bacteria to perceive many signals with intensities of the RcsCDB phosphorelay previously adjusted by IgaA. This idea of a correspondence between RcsB phosphorylation and IgaA function is now testable, for example by measuring phosphorylation status of RcsB in the different cysteine variants of IgaA reported here. We should also not discard other factors contributing to the role that IgaA has in shaping the RcsB regulon.

Another point of interest raised by our data accounts for the interaction recently reported for RcsF and IgaA as an event triggering the activation of the RcsCDB phosphorelay (Cho et al., 2014). Disulfide bonds are favored in oxidative environments as the periplasm (Denoncin and Collet, 2013; Goemans et al., 2014), and it is in this environment where the hypothetical RcsF-IgaA interaction takes place. Whether this interaction is influenced by the conserved cysteines of each partner is unknown and future studies directed to test the effect of cysteine mutations or the presence/absence of the C404–C425 disulfide bond could be therefore of much interest.

## REFERENCES

- Bittner, L. M., Westphal, K., and Narberhaus, F. (2015). Conditional proteolysis of the membrane protein Yfgm by the Ftsh protease depends on a novel n-terminal degron. *J. Biol. Chem.* 290, 19367–19378. doi: 10.1074/jbc.M115.648550
- Bouwman, C. W., Kohli, M., Killoran, A., Touchie, G. A., Kadner, R. J., and Martin, N. L. (2003). Characterization of SrgA, a *Salmonella enterica* serovar typhimurium virulence plasmid-encoded paralogue of the disulfide oxidoreductase dsba, essential for biogenesis of plasmid-encoded fimbriae. *J. Bacteriol.* 185, 991–1000. doi: 10.1128/JB.185.3.991-1000.2003
- Cano, D. A., Dominguez-Bernal, G., Tierrez, A., Garcia-Del Portillo, F., and Casadesus, J. (2002). Regulation of capsule synthesis and cell motility in *Salmonella enterica* by the essential gene *igaA*. *Genetics* 162, 1513–1523.
- Cano, D. A., Martinez-Moya, M., Pucciarelli, M. G., Groisman, E. A., Casadesus, J., and Garcia-Del Portillo, F. (2001). *Salmonella enterica* serovar Typhimurium response involved in attenuation of pathogen intracellular proliferation. *Infect. Immun.* 69, 6463–6474. doi: 10.1128/IAI.69.10.6463-6474.2001
- Casino, P., Miguel-Romero, L., Huesa, J., Garcia, P., Garcia-Del Portillo, F., and Marina, A. (2017). Conformational dynamism for DNA interaction in

## CONCLUSION

Our work provides evidence for the essential role played by one disulfide bond (C404–C425) of IgaA in its function as attenuator of the RcsCDB phosphorelay. Given the conservation of these two periplasmic cysteines in all IgaA orthologs known in enteric bacteria, the data reported here support an important structural role for this disulfide bond, probably facilitating an active conformation to the major periplasmic domain.

## AUTHOR CONTRIBUTIONS

Experimental design, methodology and investigation: MP and LR. Conceptualization and supervision: FG-dP. Writing – original draft: MP, LR, and FG-dP. Writing – reviewing and editing: FG-dP.

## FUNDING

Work in our laboratory is supported by grants BIO2016-77639-P (AEI/FEDER, UE) and PCIN-2016-082 (to FG-dP) from the Spanish Ministry of Economy and Competitiveness and European Regional Development Funds (FEDER).

## ACKNOWLEDGMENT

We thank Juan J. Cestero (Centro Nacional de Biotecnología, CNB-CSIC) for the construction of the  $\Delta dsbA$  mutant.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02605/full#supplementary-material>

- the *Salmonella* RcsB response regulator. *Nucleic Acids Res.* doi: 10.1093/nar/gkx1164 [Epub ahead of print].
- Cho, S. H., Szewczyk, J., Pesavento, C., Zietek, M., Banzhaf, M., Roszczenko, P., et al. (2014). Detecting envelope stress by monitoring beta-barrel assembly. *Cell* 159, 1652–1664. doi: 10.1016/j.cell.2014.11.045
- Denoncin, K., and Collet, J. F. (2013). Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead. *Antioxid. Redox Signal.* 19, 63–71. doi: 10.1089/ars.2012.4864
- Denoncin, K., Nicolaes, V., Cho, S. H., Leverrier, P., and Collet, J. F. (2013). Protein disulfide bond formation in the periplasm: determination of the in vivo redox state of cysteine residues. *Methods Mol. Biol.* 966, 325–336. doi: 10.1007/978-1-62703-245-2\_20
- Dierksen, K. P., and Trempey, J. E. (1996). Identification of a second RcsA protein, a positive regulator of colanic acid capsular polysaccharide genes, in *Escherichia coli*. *J. Bacteriol.* 178, 5053–5056. doi: 10.1128/jb.178.16.5053-5056.1996
- Dominguez-Bernal, G., Pucciarelli, M. G., Ramos-Morales, F., Garcia-Quintanilla, M., Cano, D. A., Casadesus, J., et al. (2004). Repression of the RcsC-YojN-RcsB phosphorelay by the IgaA protein is a requisite for *Salmonella* virulence. *Mol. Microbiol.* 53, 1437–1449. doi: 10.1111/j.1365-2958.2004.04213.x

- Donnenberg, M. S., and Kaper, J. B. (1991). Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect. Immun.* 59, 4310–4317.
- Evans, K. L., Kannan, S., Li, G., De Pedro, M. A., and Young, K. D. (2013). Eliminating a set of four penicillin binding proteins triggers the Rcs phosphorelay and Cpx stress responses in *Escherichia coli*. *J. Bacteriol.* 195, 4415–4424. doi: 10.1128/JB.00596-13
- Farris, C., Sanowar, S., Bader, M. W., Pfuetschner, R., and Miller, S. I. (2010). Antimicrobial peptides activate the Rcs regulon through the outer membrane lipoprotein RcsF. *J. Bacteriol.* 192, 4894–4903. doi: 10.1128/JB.00505-10
- Gao, R., and Stock, A. M. (2017). Quantitative kinetic analyses of shutting off a two-component system. *mBio* 8:e00412-17. doi: 10.1128/mBio.00412-17
- Goemans, C., Denoncin, K., and Collet, J. F. (2014). Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta* 1843, 1517–1528. doi: 10.1016/j.bbamcr.2013.10.014
- Hagiwara, D., Sugiura, M., Oshima, T., Mori, H., Aiba, H., Yamashino, T., et al. (2003). Genome-wide analyses revealing a signaling network of the RcsC-YojN-RcsB phosphorelay system in *Escherichia coli*. *J. Bacteriol.* 185, 5735–5746. doi: 10.1128/JB.185.19.5735-5746.2003
- Howery, K. E., Clemmer, K. M., and Rather, P. N. (2016). The Rcs regulon in *Proteus mirabilis*: implications for motility, biofilm formation, and virulence. *Curr. Genet.* 62, 775–789. doi: 10.1007/s00294-016-0579-1
- Jurado, P., De Lorenzo, V., and Fernandez, L. A. (2006). Thioredoxin fusions increase folding of single chain Fv antibodies in the cytoplasm of *Escherichia coli*: evidence that chaperone activity is the prime effect of thioredoxin. *J. Mol. Biol.* 357, 49–61. doi: 10.1016/j.jmb.2005.12.058
- Konovalova, A., Mitchell, A. M., and Silhavy, T. J. (2016). A lipoprotein/beta-barrel complex monitors lipopolysaccharide integrity transducing information across the outer membrane. *Elife* 5:e15276. doi: 10.7554/eLife.15276
- Latasa, C., Garcia, B., Echeverez, M., Toledo-Arana, A., Valle, J., Campoy, S., et al. (2012). *Salmonella* biofilm development depends on the phosphorylation status of RcsB. *J. Bacteriol.* 194, 3708–3722. doi: 10.1128/JB.00361-12
- Leverrier, P., Declercq, J. P., Denoncin, K., Vertommen, D., Hiniker, A., Cho, S. H., et al. (2011). Crystal structure of the outer membrane protein RcsF, a new substrate for the periplasmic protein-disulfide isomerase DsbC. *J. Biol. Chem.* 286, 16734–16742. doi: 10.1074/jbc.M111.224865
- Lin, D., Kim, B., and Schlauch, J. M. (2009). DsbL and DsbI contribute to periplasmic disulfide bond formation in *Salmonella enterica* serovar Typhimurium. *Microbiology* 155, 4014–4024. doi: 10.1099/mic.0.032904-0
- Majdalani, N., and Gottesman, S. (2005). The Rcs phosphorelay: a complex signal transduction system. *Annu. Rev. Microbiol.* 59, 379–405. doi: 10.1146/annurev.micro.59.050405.101230
- Majdalani, N., and Gottesman, S. (2007). Genetic dissection of signaling through the Rcs phosphorelay. *Methods Enzymol.* 423, 349–362. doi: 10.1016/S0076-6879(07)23016-2
- Majdalani, N., Heck, M., Stout, V., and Gottesman, S. (2005). Role of RcsF in signaling to the Rcs phosphorelay pathway in *Escherichia coli*. *J. Bacteriol.* 187, 6770–6778. doi: 10.1128/JB.187.19.6770-6778.2005
- Mariscotti, J. F., and Garcia-Del Portillo, F. (2008). Instability of the *Salmonella* RcsCDB signalling system in the absence of the attenuator IgaA. *Microbiology* 154, 1372–1383. doi: 10.1099/mic.0.2007/015891-0
- Mariscotti, J. F., and Garcia-del Portillo, F. (2009). Genome expression analyses revealing the modulation of the *Salmonella* Rcs regulon by the attenuator IgaA. *J. Bacteriol.* 191, 1855–1867. doi: 10.1128/JB.01604-08
- Miki, T., Okada, N., and Danbara, H. (2004). Two periplasmic disulfide oxidoreductases, DsbA and SrgA, target outer membrane protein SpiA, a component of the *Salmonella* pathogenicity island 2 type III secretion system. *J. Biol. Chem.* 279, 34631–34642. doi: 10.1074/jbc.M402760200
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Navasa, N., Rodriguez-Aparicio, L., Ferrero, M. A., Monteagudo-Mera, A., and Martinez-Blanco, H. (2013). Polysialic and colanic acids metabolism in *Escherichia coli* K92 is regulated by RcsA and RcsB. *Biosci. Rep.* 33:e00038. doi: 10.1042/BSR20130018
- Pannen, D., Fabisch, M., Gausling, L., and Schnetz, K. (2016). Interaction of the RcsB response regulator with auxiliary transcription regulators in *Escherichia coli*. *J. Biol. Chem.* 291, 2357–2370. doi: 10.1074/jbc.M115.696815
- Rogov, V. V., Rogova, N. Y., Bernhard, F., Lohr, F., and Dotsch, V. (2011). A disulfide bridge network within the soluble periplasmic domain determines structure and function of the outer membrane protein RcsF. *J. Biol. Chem.* 286, 18775–18783. doi: 10.1074/jbc.M111.230185
- Rosu, V., Chevance, F. F., Karlinsey, J. E., Hirano, T., and Hughes, K. T. (2006). Translation inhibition of the *Salmonella* *fliC* gene by the *fliC* 5' untranslated region, *fliC* coding sequences, and FlgM. *J. Bacteriol.* 188, 4497–4507. doi: 10.1128/JB.01552-05
- Westphal, K., Langklotz, S., Thomanek, N., and Narberhaus, F. (2012). A trapping approach reveals novel substrates and physiological functions of the essential protease FtsH in *Escherichia coli*. *J. Biol. Chem.* 287, 42962–42971. doi: 10.1074/jbc.M112.388470
- Wolanin, P. M., Thomason, P. A., and Stock, J. B. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol.* 3:REVIEWS3013.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pucciarelli, Rodríguez and García-del Portillo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Fluorescence Imaging of *Streptococcus pneumoniae* with the *Helix pomatia* agglutinin (HPA) As a Potential, Rapid Diagnostic Tool

Mirian Domenech<sup>1,2</sup> and Ernesto García<sup>1,2\*</sup>

<sup>1</sup> Departamento de Microbiología Molecular y Biología de las Infecciones, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Madrid, Spain

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Universiti Sultan Zainal Abidin,  
Malaysia

### Reviewed by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain  
Darío García De Viedma,  
Hospital General Universitario  
Gregorio Marañón, Spain  
Analia Rial,  
Facultad de Medicina, Universidad de  
la Republica de Uruguay, Uruguay  
Natalia Munoz Wolf,  
Trinity College, Dublin, Ireland

### \*Correspondence:

Ernesto García  
e.garcia@cib.csic.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 27 January 2017

**Accepted:** 30 June 2017

**Published:** 18 July 2017

### Citation:

Domenech M and García E (2017)  
Fluorescence Imaging of  
*Streptococcus pneumoniae* with the  
*Helix pomatia* agglutinin (HPA) As a  
Potential, Rapid Diagnostic Tool.  
Front. Microbiol. 8:1333.  
doi: 10.3389/fmicb.2017.01333

*Streptococcus pneumoniae* is a common human pathogen and a major causal agent of life-threatening infections that can either be respiratory or non-respiratory. It is well known that the *Helix pomatia* (edible snail) agglutinin (HPA) lectin shows specificity for terminal  $\alpha$ GalNAc residues present, among other locations, in the Forssman pentasaccharide ( $\alpha$ GalNAc1 $\rightarrow$ 3 $\beta$ GalNAc1 $\rightarrow$ 3 $\alpha$ Gal1 $\rightarrow$ 4 $\beta$ Gal1 $\rightarrow$ 4 $\beta$ Glc). Based on experiments involving choline-independent mutants and different growth conditions, we propose here that HPA recognizes the  $\alpha$ GalNAc terminal residues of the cell wall teichoic and lipoteichoic acids of *S. pneumoniae*. In addition, experimental evidence showing that pneumococci can be specifically labeled with HPA when growing as planktonic cultures as well as in mixed biofilms of *S. pneumoniae* and *Haemophilus influenzae* has been obtained. It should be underlined that pneumococci were HPA-labeled despite of the presence of a capsule. Although some non-pneumococcal species also bind the agglutinin, HPA-binding combined with fluorescence microscopy constitutes a suitable tool for identifying *S. pneumoniae* and, if used in conjunction with Gram staining and/or other suitable technique like antigen detection, it may potentially facilitate a fast and accurate diagnosis of pneumococcal infections.

**Keywords:** *Streptococcus pneumoniae*, Forssman antigen, binding lectins, teichoic acids, fluorescence microscopy

## INTRODUCTION

*Streptococcus pneumoniae*, the pneumococcus, is a leading human pathogen and one of the foremost etiologic agents of invasive diseases such as bacteremic community-acquired pneumonia (CAP), bacteremia, and meningitis, mainly in children, the elderly, and immunocompromised patients. Indeed, the pneumococcus is one of the major causes of non-invasive diseases such as non-bacteremic CAP, acute otitis media, sinusitis and conjunctivitis. In 2015, CAP accounted for 16% of all deaths of children under 5 years old globally and 920,000 deaths globally in children of all ages (World Health Organization, 2016). *S. pneumoniae* is the commonest bacterial causative agent across all ages, accounting for 30–40% of CAP cases (Haq et al., 2017). Other bacterial causes of CAP include *Streptococcus pyogenes* (group A streptococci; GAS), and *Streptococcus agalactiae* (group B streptococci; GBS) in infants. *Staphylococcus aureus* is associated with round pneumonia, a well-defined round area of consolidation visible on chest radiographs. Despite the current vaccination



programs, *Haemophilus influenzae* remains prevalent in several developing countries (Hajjeh et al., 2013) and *Mycoplasma pneumoniae* accounts for up to a third of all cases and is a common cause of atypical CAP.

Since mortality rates are highest during the first days of admission (Ewig et al., 2009), early diagnosis and treatment may have a crucial role in curing the patient or in reducing their morbidity and mortality, particularly in the era of antibiotic resistance (Vila et al., 2017). Rapid identification techniques are even more important in bacterial meningitis patients, since delayed initiation of antibiotic treatment is strongly associated with death and poor outcome; as a consequence, it has been recently recommended that antibiotic treatment in these patients should be started as soon as possible, and the time period from entering the hospital to initiation of antibiotic treatment should not exceed 1 h (van de Beek et al., 2016). In every case, the gold standard diagnostic method is still culture. Unfortunately, this is time consuming (24–72 h) and includes the inoculation of appropriate media, subculturing, and phenotype-based characterization via biochemical testing, along with antibiotic susceptibility testing. Currently, bacteriological diagnosis is making progress in molecular biology techniques such as PCR, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), and whole genome sequencing (Clark, 2015; Pai et al., 2015; McGill et al., 2016; Torres et al., 2016). These techniques have been developed to decrease the time for the initiation of an optimal antimicrobial therapy for improving clinical outcomes. Unfortunately, although microbiological diagnosis is very important for a correct clinical management of the disease—particularly because spreading of antibiotic multiresistance is an increasing global concern—, microbiological identification is lacking in many instances approaching 50% in CAP cases (Cilloniz et al., 2016).

Most bacteria in nature exist in multispecies communities known as biofilms. Multispecies biofilms are structured and spatially defined communities where species interact both intra- and interspecifically (Røder et al., 2016). Imaging techniques are useful for identifying multiple species, which provides information on their spatial organization. Fluorescence *in situ* hybridization (FISH) and many advanced versions of the FISH technique have been implemented for different purposes; these are well-established means of visualizing and identifying microbial groups or species in natural and artificially created multispecies biofilms (Amann and Fuchs, 2008; Liu et al., 2016; Røder et al., 2016). Although FISH is typically time-consuming and destructive to the samples, it is however possible to visualize different species simultaneously (Schimak et al., 2016). Fluorescence imaging of individual species can also be achieved by genetically marking the species with genes coding for different fluorescent or bioluminescent proteins (Kjos et al., 2015). Unfortunately, not all bacteria can be fluorescently tagged, particularly those from natural samples.

Polymicrobial biofilms are abundant in clinical diseases such as acute otitis media, a significant public health problem worldwide (Monasta et al., 2012), particularly among children (Ahmed et al., 2014). Acute otitis media is preceded by

the nasopharyngeal carriage of bacterial pathogens like *S. pneumoniae*, non-typeable (NT) *H. influenzae*, and *Moraxella catarrhalis* (Ngo et al., 2016). Moreover, the concurrent carriage of these pathogens is a predictor of clinical pneumonia (Chochua et al., 2016). When growing in an *in vitro* mixed biofilm, *S. pneumoniae* and NT *H. influenzae* cells appear to accomplish a strong inter-population cooperation, i.e., metabolic interdependence or mutualism (Momeni et al., 2013), as indicated by the finding that pneumococci were intermixed with NT *H. influenzae* cells throughout the biofilm (Domenech and García, 2017). This was determined using the *Helix pomatia* (edible snail) agglutinin (HPA) that unambiguously stained *S. pneumoniae* cells in the mixed biofilms. The binding preference of HPA has been reported to be the Forssman pentasaccharide ( $\alpha\text{GalNAc1}\rightarrow3\beta\text{GalNAc1}\rightarrow3\alpha\text{Gal1}\rightarrow4\beta\text{Gal1}\rightarrow4\beta\text{Glc}$ ) > blood group A substance ( $\alpha\text{GalNAc1}\rightarrow3[\alpha\text{Fuc1}\rightarrow2]\text{Gal}$ ) > Tn antigen ( $\alpha\text{GalNAc-Ser/Thr}$ ) > GalNAc > GlcNAc, confirming its specificity for terminal  $\alpha\text{GalNAc}$  residues (Wu and Sugii, 1991; Cooling, 2015). The Forssman antigen (FA) can be defined as a substance that provokes the appearance of sheep red blood cell hemolytic antibodies when injected into rabbits; it is a glycolipid with the structure  $\text{GalNAc}\alpha1\rightarrow3\text{GalNAc}\beta1\rightarrow3\text{Gal}\alpha1\rightarrow4\text{Gal}\beta1\rightarrow4\text{Glc}\beta1\rightarrow1\text{Cer}$  (Siddiqui and Hakomori, 1971). Originally found in the tissues of different animals (although not humans; Yamamoto et al., 2012), FA was subsequently discovered in some bacteria, including *S. pneumoniae* (Jenkin, 1963). It has been shown that the Forssman cross-reactive material of *S. pneumoniae* is the type IV, ribitol phosphate-containing, membrane anchored lipoteichoic acid (pnLTA) (Briles and Tomasz, 1973; Gisch et al., 2013). The non-lipid terminus of pnLTA consists of 6-O-PCho- $\alpha$ -D-GalpNAc-(1 $\rightarrow$ 3)-6-O-PCho- $\beta$ -D-GalpNAc (Seo et al., 2008); this disaccharide represents a structural feature that is able to partly explain the FA properties of pnLTA (Gisch et al., 2013). In addition and unlike other bacterial species, pnLTA and the *S. pneumoniae* peptidoglycan-bound teichoic acid (pnWTA) have identical chain structures (for a recent review, see reference by Gisch et al. (2015a)). Based on this information it can be assumed that HPA binds to both WTA and LTA of *S. pneumoniae*.

In the present study we report that HPA binds to the teichoic acids (TA) of encapsulated and non-encapsulated pneumococcal cells grown either planktonically or forming biofilms. In addition to *S. pneumoniae*, HPA labeling of other bacterial species, e.g., some *S. aureus* strains, has also been observed. We propose that, in combination with other widespread rapid techniques, labeling with HPA in biological fluids may represent a helpful technique for the fast and accurate diagnosis of pneumococcal diseases.

## MATERIALS AND METHODS

### Bacteria and Growth Conditions

The bacteria used in this study are listed in Table 1. Streptococci, staphylococci, enterococci, and *Pseudomonas aeruginosa* were grown in Todd-Hewitt broth supplemented with 0.5% yeast extract (THY). For planktonic growth, the NT *H. influenzae* strain 54997 was incubated in brain heart infusion (BHI)

**TABLE 1** | Bacterial strains used in this study.

Strain	Description	Source or References <sup>a</sup>	HPA labeling <sup>b</sup>
<i>S. pneumoniae</i>			
D39	Serotype 2; clinical isolate	Lanie et al., 2007	+
P007	Serotype 3; laboratory transformant	Domenech et al., 2009	+
R6	Non-encapsulated D39 derivative	Hoskins et al., 2001	+
MNZ67	Non-encapsulated; clinical isolate	Park et al., 2012	+
1001	Serotype 3; clinical isolate	J. Liñares <sup>c</sup>	+
2951/13	Serotype 12F; clinical isolate	Domenech et al., 2015	+
Tupelo_VT	Serotype 14; clinical isolate	Moscoso et al., 2010	+
1064	Serotype 19F; clinical isolate	A. Fenoll <sup>d</sup>	+
1178/17	Serotype 22F; clinical isolate	L. Vicioso <sup>d</sup>	+
2971/13	Serotype 23B; clinical isolate	Domenech et al., 2015	+
JY2190	Choline-independent Rx1 mutant; non-encapsulated	Yother et al., 1998	+
P501	Choline-independent R6 mutant; non-encapsulated	González et al., 2008	+
<i>Streptococcus pseudopneumoniae</i>			
CCUG 49455	Type strain	CCUG	–
<i>Streptococcus mitis</i>			
NCTC 12261	Type strain	NCTC	–
SK137	Biovar 1; strain with choline-containing teichoic acids	Bergström et al., 2000	+
SK598	Biovar 1; strain with ethanolamine-containing teichoic acids	Bergström et al., 2003	–
<i>Streptococcus oralis</i>			
NCTC 11427	Type strain	Kilpper-Bälz et al., 1985	–
<i>Streptococcus pyogenes</i>			
CECT 985	Type strain; Lancefield group A	CECT	–
<i>Streptococcus agalactiae</i>			
CECT 183	Type strain; Lancefield group B	CECT	–
<i>Streptococcus dysgalactiae</i> subsp. <i>equisimilis</i>			
CECT 926	ATCC 9542; Lancefield group C	CECT	+
<i>Streptococcus suis</i>			
298	Serotype 9; Lancefield group D <sup>e</sup>	A. I. Vela <sup>f</sup>	+/-
357	Serotype 2; Lancefield group D	A. I. Vela	+/-
<i>Staphylococcus aureus</i>			
CECT 86	Type strain	Shiroma et al., 2015; CECT	+
15981	Clinical isolate; good biofilm former	Valle et al., 2003	+
<i>Staphylococcus epidermidis</i>			
CECT 231	Good biofilm former	Domenech et al., 2011; CECT	–
<i>Enterococcus faecalis</i>			
CECT 481	Type strain	CECT	–
<i>Haemophilus influenzae</i>			
54997	Nontypeable	García-Cobos et al., 2014	–
<i>Pseudomonas aeruginosa</i>			
PAO1	ATCC 15692	ATCC	–

<sup>a</sup> ATCC, American Type Culture Collection; CECT, Colección Española de Cultivos Tipo; CCUG, Culture Collection, University of Göteborg; NCTC, National Collection of Type Cultures.

<sup>b</sup> +, positive; –, negative; +/-, slightly positive.

<sup>c</sup> Instituto de Investigación Biomédica de Bellvitge (IDIBEL); Barcelona (Spain).

<sup>d</sup> Laboratorio de Referencia de Neumococos; Centro Nacional de Microbiología (CNM-ISCIII); Majadahonda (Madrid; Spain).

<sup>e</sup> Although early studies reported that some strains of *S. suis* contained streptococcal group D antigen, more recent results indicated that the group R and group D antigens were similar and crossreacted. To date, the species belonging to the *Streptococcus bovis* group constitute the non-enterococcal group D streptococci (Dekker and Lau, 2016).

<sup>f</sup> Facultad de Veterinaria; Universidad Complutense de Madrid; Madrid (Spain).

supplemented with haemin and NAD (15 µg/ml each) (sBHI). In some experiments, *S. pneumoniae* was grown in C medium (Lacks and Hotchkiss, 1960) supplemented with 0.08% yeast extract (Difco Laboratories; C+Y medium) or a chemically defined medium (Cden) supplemented (or not) with either choline (5 µg/ml) (Cden-choline) or ethanolamine (40 µg/ml) (Cden-EA) (Tomasz, 1968). Cells were incubated at 37°C without shaking. Bacterial growth was monitored by measuring the absorbance at 595 nm ( $A_{595}$ ).

Formation of mixed biofilms of *S. pneumoniae* and NT *H. influenzae* was performed as described elsewhere (Domenech and García, 2017). Briefly, cultures of *S. pneumoniae* strain R6 and *H. influenzae* strain 54997 were grown to mid-exponential phase in C+Y medium supplemented with haemin and NAD (15 µg/ml each) [s(C+Y)], diluted to  $\approx 5 \times 10^6$  colony-forming units (cfu)/ml and mixed in an 1:1 proportion. Two milliliter of the mixtures were then distributed into the wells of a glass-bottomed dish (WillCo-dish, WillCo Wells B. V., The Netherlands) and incubated for 6 h at 37°C under 5% CO<sub>2</sub>. For species biofilm formation, 2 ml of the individual cultures ( $5 \times 10^6$  cfu/ml each) were independently inoculated as indicated above for mixed biofilms.

All studies which involved the handling of virulent bacteria, whole blood, or blood derivatives were undertaken at the biosafety level II laboratory of Centro de Investigaciones Biológicas. It should be mentioned that, according to the supplier (Innovative Research), the whole human blood used had been tested by FDA-approved methods for human immunodeficiency virus RNA, antibodies to immunodeficiency virus, antibodies to hepatitis C virus, hepatitis C virus RNA, hepatitis B virus, hepatitis B surface antigen, and syphilis.

## Staining with HPA Lectin

Exponentially growing cells of various bacterial species were centrifuged, washed and suspended in phosphate-buffered saline (PBS). After incubation for 15 min at room temperature in the dark with HPA lectin conjugated to Alexa Fluor-488 (2.5–25 µg/ml), cells were centrifuged again and resuspended in PBS. Bacteria were observed under a Leica DM4000B fluorescence microscope equipped with an L5 filter (bandpass 480/40), and viewed under a Leica HCX PL FLUOTAR 40×/0.75 objective or an HC PL APO 63×/1.40–0.60 oil objective. In early experiments, bacteria were also diluted into fetal bovine serum (from Sigma-Aldrich), or into defibrinated sheep blood (from Biomedics or Oxoid). Afterwards, experiments were also carried out with groups A and O citrated human whole blood (from Innovative Research). It should be underlined that since HPA labeling does not require the presence of divalent cations (Kobayashi et al., 2014), blood treated to prevent its coagulation (e.g., citrate- or EDTA-treated) can be used.

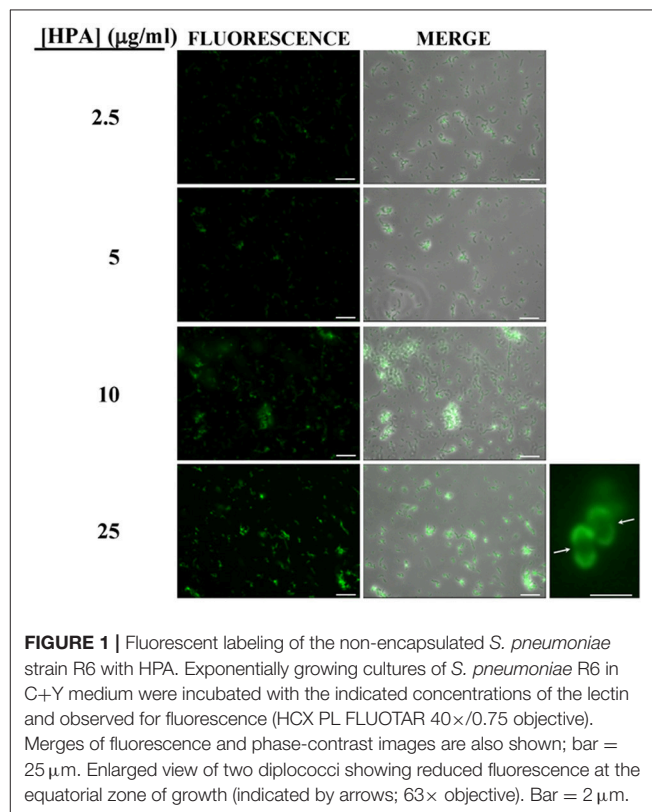
For biofilm observation, the culture medium was removed and the biofilm rinsed with sterile water to remove non-adherent bacteria. Staining was performed with HPA and SYTO 59 and biofilms were gently rinsed with PBS. Observations were made using a Leica TCS-SP2-AOBS-UV confocal laser scanning microscope (CLSM) equipped with an argon ion laser. Images were analyzed using LCS software from Leica. Projections were

obtained in the planes  $x-y$  (individual scans at 0.5 µm intervals) and  $x-z$  (images at 5 µm intervals).

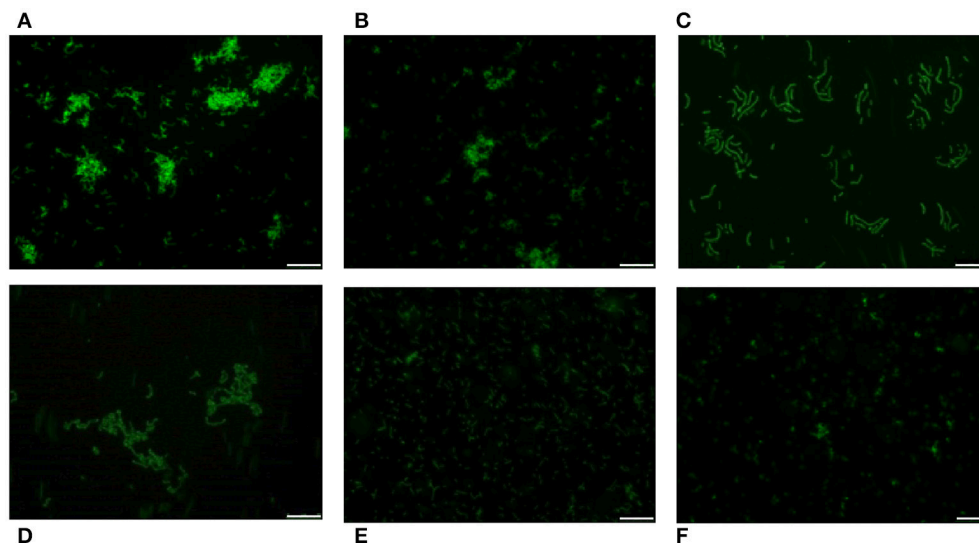
## RESULTS

### HPA Binding by Different Bacteria

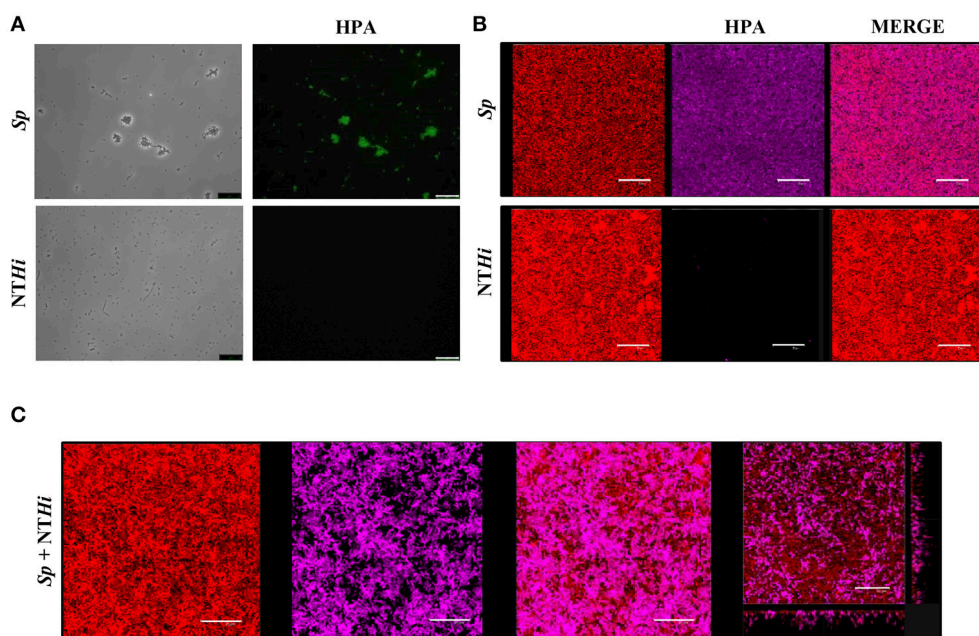
Pneumococci were clearly labeled with HPA when this was present in concentrations of 2.5–25 µg/ml (**Figure 1**). Closer examination showed that HPA labeling was not uniform across the pneumococcal surface; reduced fluorescence was noticed in the equatorial zone of growth, the place where new cell wall material is incorporated (Gisch et al., 2015a). The ability to bind HPA was not exclusive to pneumococci; the surfaces of other bacteria too were labeled with the lectin, i.e., *Streptococcus mitis* SK137 and some strains of *Streptococcus dysgalactiae* subsp. *equisimilis*, *Streptococcus suis* and *S. aureus* (**Figure 2**, **Table 1**). Notably, HPA was unable to label the type strains (<sup>T</sup>) of *Streptococcus pseudopneumoniae*, *S. mitis*, or *Streptococcus oralis*, the three closest relatives of the pneumococcus. Notably, cells of other relevant pathogens, either Gram-positive (GAS, GBS, *Enterococcus faecalis* and *Staphylococcus epidermidis*) or Gram-negative (*H. influenzae* and *P. aeruginosa*), did not bind the lectin (**Table 1**). Besides, and in agreement with previous results (Domenech and García, 2017), *S. pneumoniae* (but not NT *H. influenzae*) were also labeled with HPA when growing as mixed biofilms (**Figure 3**).



**FIGURE 1 |** Fluorescent labeling of the non-encapsulated *S. pneumoniae* strain R6 with HPA. Exponentially growing cultures of *S. pneumoniae* R6 in C+Y medium were incubated with the indicated concentrations of the lectin and observed for fluorescence (HCX PL FLUOTAR 40×/0.75 objective). Merges of fluorescence and phase-contrast images are also shown; bar = 25 µm. Enlarged view of two diplococci showing reduced fluorescence at the equatorial zone of growth (indicated by arrows; 63× objective). Bar = 2 µm.



**FIGURE 2 |** HPA labeling of various streptococci in C+Y medium. First row: *S. pneumoniae* strains R6 (non-encapsulated) (A), D39 (serotype 2) (B), and P007 (serotype 3) (C). Second row: *S. suis* 298 (D), *S. dysgalactiae* subsp. *equisimilis* CECT 926 (E), and *S. aureus* type strain (F). Bar = 25  $\mu$ m.

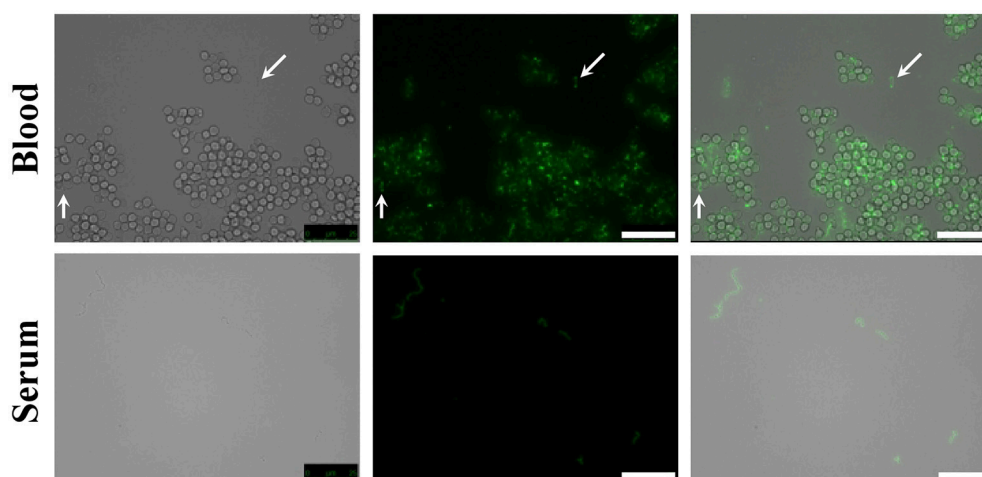


**FIGURE 3 |** *S. pneumoniae* labeling with HPA in mixed biofilms. (A) Fluorescent labeling of the *S. pneumoniae* (Sp) strain R6 and the non-typeable *H. influenzae* (NTHi) 54997 with HPA. Biofilms formed by Sp R6 or NTHi 54997 (B), or both pathogens (C) were stained with a combination of SYTO 59 (red) and HPA conjugated to Alexa Fluor-488 (pink). An orthogonal projection of a CLSM image showing a representative region of the x-y plane over the depth of the biofilm in both x-z and y-z dimensions of the mixed biofilm is also shown at the bottom right part of the figure. Planktonic and biofilms cultures were incubated with HPA (25  $\mu$ g/ml). Bar = 25  $\mu$ m.

The presence of the capsule does not appear to hinder HPA binding to the *S. pneumoniae* surface since strains D39 (serotype 2) and P007 (a heavily encapsulated serotype 3 transformant) were efficiently labeled (Figures 2B,C). Cells of *S. pneumoniae* D39 were also positive for HPA binding when diluted to  $2.5 \times 10^6$  cfu/ml into defibrinated sheep blood or

into fetal bovine serum, although excess fluorescence on the erythrocyte surface may partly hinder a distinct pneumococcal identification (Figure 4). As HPA also binds to the human blood group A antigen (Matsui et al., 2001), that interference, which is due to the fact that sheep erythrocytes harbor FA on their surface (see above), should disappear using group non-A





**FIGURE 4 |** HPA labeling of *S. pneumoniae* in non-human blood and serum. The type 2 encapsulated *S. pneumoniae* strain D39 was diluted to  $2.5 \times 10^6$  cfu/ml into sheep blood or fetal bovine serum ( $63\times$  objective). Cultures were incubated with HPA ( $25\ \mu\text{g/ml}$ ). Arrows point to pneumococcal cells. Bar =  $25\ \mu\text{m}$ .

(e.g., group O) human whole blood. In addition, it should be mentioned that up to 98 pneumococcal capsular polysaccharides differing in sugar composition and linkages have been described to date (Geno et al., 2017). Besides, invasive pneumococcal disease and pneumonia rates have decreased in most countries following the introduction of conjugate pneumococcal vaccines (PCVs). However, after PCV implementation, current data show that more non-vaccine serotypes increased in frequency than decreased, which is consistent with vaccine-induced replacement. Clinical pneumococcal isolates of six different serotypes—including three emerging serotypes that are currently among the predominant non-PCV13 serotypes worldwide, i.e., serotypes 12F, 22F, and 23B; Balsells et al., 2017—were diluted to about  $2.5 \times 10^6$  cfu/ml into whole human blood, incubated with HPA, and observed under the microscope. Moreover, due to the increasing clinical importance of non-encapsulated *S. pneumoniae* (Keller et al., 2016), a representative strain (MNZ67) was also investigated. All the pneumococcal isolates tested bound the lectin and, as expected, HPA-labeled pneumococci were particularly noticeable using group O human blood (Figure 5). It is worth mentioning that any possible interference caused by an excess fluorescence on the surface of group A human erythrocytes could be virtually abolished by partly sedimenting the blood cells at low speed ( $1,000 \times g$ ; 1 min; room temperature) before HPA addition (Figure 5A, bottom right).

## HPA Labeling of Pneumococcal Teichoic Acids

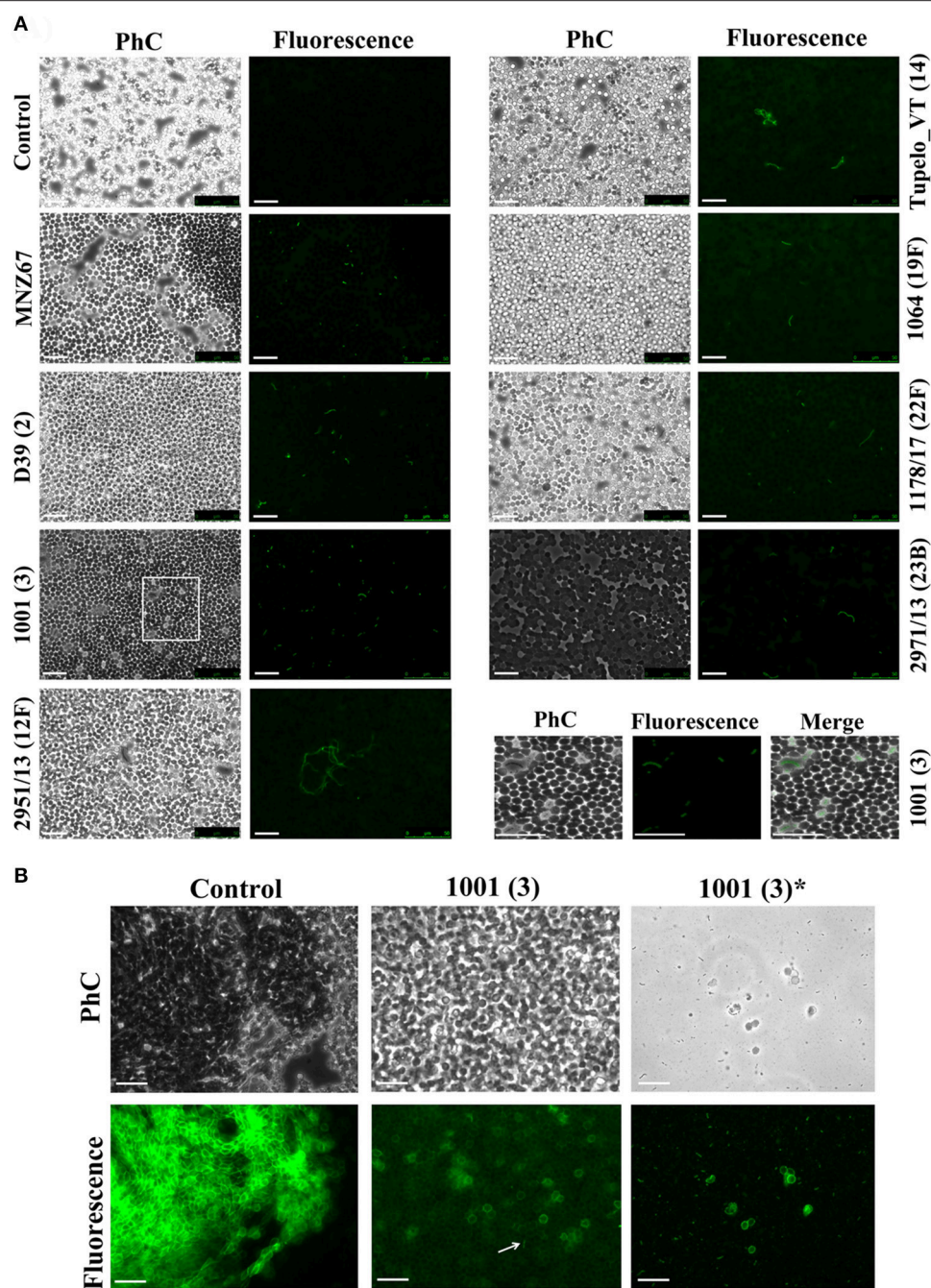
With the possible exception of *S. pseudopneumoniae* (González et al., 2008), the nutritional requirement of pneumococci for the amino alcohol choline (as a component of pnWTA and pnLTA) appears to be an exclusive trait (Rane and Subbarow, 1940). Nevertheless, several choline-independent pneumococcal mutants have been characterized in the last years, and it has been suggested that the absence of choline incorporation might affect

the structure of TAs as well as the composition of the cell wall. Actually, it has been shown that pnLTA and pnWTA isolated from these mutants were free of phosphocholine and other phosphorylated aminoalcohols (Yother et al., 1998). However, and as deduced from the *in vivo* cell labeling with HPA of two double *tacF* mutants that form long chains of cells (and are autolysis-defective) when grown in media lacking any amino alcohol (strains JY2190 and P501), the absence of choline residues in TAs does not appear to modify HPA binding by *S. pneumoniae* (Figure 6).

It is well known that pneumococci growing in Cden medium containing EA instead of choline form long chains, do not autolyze, and are non-transformable (Tomasz, 1968). In the present work, the EA-grown cells were unable to bind HPA (Figures 7A,D). The same was seen when *S. pneumoniae* R6 was incubated in C+Y supplemented with 2% choline chloride (data not shown). It has been reported that upon the addition of choline ( $5\ \mu\text{g/ml}$ ) to EA-grown pneumococci, these cells revert to the normal phenotype (Tomasz et al., 1975). As expected, EA-grown cells became entirely HPA-labeled 180 min ( $\approx 4$  generations) after shifting to a choline-containing medium (Figures 7B,E). As an alternative and complementary model, we examined the *S. mitis* strain SK598, which is unique in that its WTA and LTA contain EA instead of choline even when incubated in a choline-containing medium (Bergström et al., 2003). As observed for EA-grown *S. pneumoniae* cells, *S. mitis* SK598 was unable to bind HPA (Figures 7C,F).

## DISCUSSION

In a previous work we have shown that, in contrast to *S. pneumoniae*, NT *H. influenzae* strains did not bind HPA; thus fluorescent staining of *S. pneumoniae* with the HPA lectin revealed that pneumococci were evenly distributed throughout the *in vitro* biofilm and interspersed with NT *H. influenzae*



**FIGURE 5 |** HPA labeling of clinical pneumococcal isolates in human blood. The *S. pneumoniae* strains were diluted to  $2.5 \times 10^6$  cfu/ml into group O **(A)** and A **(B)** human blood (40 $\times$  objective). Serotypes are indicated in parentheses. Strain MNZ67 is a non-encapsulated clinical isolate. In **(A)**, the images shown at the bottom right are enlarged visions of the area marked with a rectangle in strain 1001. Control denotes non-infected blood samples. In **(B)**, those labeled with an asterisk correspond to views of the supernatant of infected group A blood that had been centrifuged (1,000  $\times$  g; 1 min; room temperature) before HPA labeling. Samples were incubated with HPA (25  $\mu$ g/ml). The arrow points to a pneumococcal diplococcus. Phc, phase-contrast micrograph; Bar = 25  $\mu$ m.

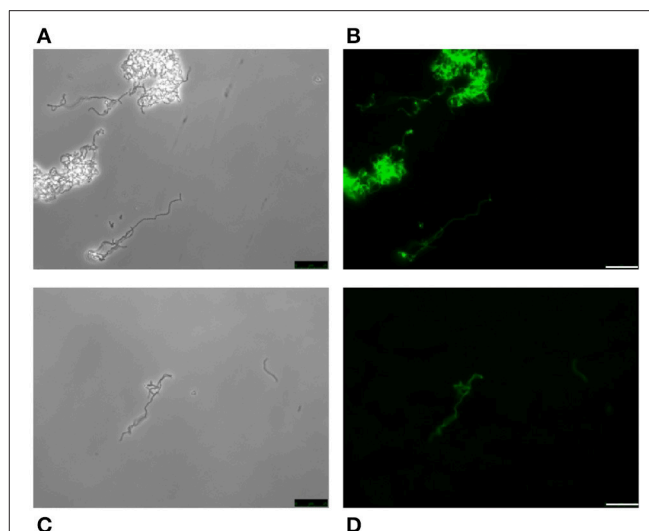
(Domenech and García, 2017). We have shown here that HPA recognition of the *S. pneumoniae* surface does not require the presence of choline residues in TAs. This observation is in agreement with the results of Gisch et al. who recently reported that the presence/absence of phosphorylcholine in the

FA terminus of pnLTA to have no effect on detection by an anti-Forssman antibody (Gisch et al., 2013). This is relevant since the number of repeating units and the phosphocholine content per repeat (mono-substituted or di-substituted) for pneumococcal TAs slightly vary among strains (Gisch et al., 2015a). It is

interesting, however, that, in the present work, EA-grown cells or *S. pneumoniae* R6 incubated in C+Y supplemented with 2% choline chloride did not bind HPA. Incubation in high choline concentrations inhibits cell separation, leaving pneumococci to grow in long chains of cells, just like EA-grown cells (Briese and Hakenbeck, 1984; Giudicelli and Tomasz, 1984). This phenotype is, at least partly, the result of the inhibition of LytB—a member of the family of choline-binding proteins (CBPs)—and the release of this and other CBPs to the medium (López and García, 2004). Both processes would be expected to occur in choline-independent strains growing in the absence of any amino alcohol (see above), although when this was performed in the present work HPA-labeling was unaffected. Interestingly, in an early study, Briles and Tomasz reported the yield of heterophile (sheep hemolytic) antibodies elicited by pneumococci to be at least 10-fold greater in choline- than in EA-containing media (Briles and Tomasz, 1975). Moreover, these authors reported that pneumococci growing in C+Y medium (choline-containing) elicit antibodies which bind poorly to EA-grown bacteria, whereas the latter elicit antibodies which bind well to choline-grown cells. The reasons for these unexpected results are still unclear, although the possibility of the existence of a choline-dependent regulatory pathway for the synthesis of pneumococcal TAs warrants future research.

To our knowledge, the concentration of free choline (or EA) in human lungs has not been reported so far. However, the normal levels of free choline and EA in human fluids are quite similar, e.g., about 1  $\mu\text{M}$  each in serum (Forteschi et al., 2016; Dereziński et al., 2017) and 2  $\mu\text{M}$  for choline or 9–15  $\mu\text{M}$  for EA in cerebrospinal fluid (CSF) (Kruse et al., 1985; Frölich et al., 1998; Ogawa et al., 2015). These data together with the early findings that choline is an effective inhibitor of the cellular incorporation of EA: addition of as little as 0.1  $\mu\text{g/ml}$  choline to a culture growing in the presence of 40  $\mu\text{g/ml}$  EA immediately inhibited further EA incorporation, and that choline was incorporated by such cultures without any detectable lag (Tomasz, 1968), strongly suggest that EA-grown cells (or pneumococci grown in high choline concentrations) are not expected to be found in nature, although they represent important model systems for *in vitro* studies.

HPA labeling allows the recognition of *S. pneumoniae* cells among a variety of other species. However, it is clear that HPA specificity is not restricted to pneumococci (Table 1). Interestingly, neither *S. pseudopneumoniae*<sup>T</sup>, *S. mitis*<sup>T</sup> nor *S. oralis*<sup>T</sup> binds the lectin; this is of note since, although these three species are very closely related to *S. pneumoniae*, monoclonal antibodies directed against the backbone and the phosphocholine residues of TAs react only with some strains of these three species (Kilian et al., 2008). The binding of HPA to these bacteria could, therefore, be mostly strain-specific. For example, in contrast to that observed for *S. mitis*<sup>T</sup>, *S. mitis* SK137 was susceptible to HPA labeling (Table 1). This was not unexpected since this particular strain has choline-containing TAs with a carbohydrate backbone identical to that of pnWTA/pnLTA, which forms the Lancefield group O antigen (Bergström et al., 2000). Quite unexpectedly, the SK137 strain only showed an average 67.1% nucleotide similarity to the



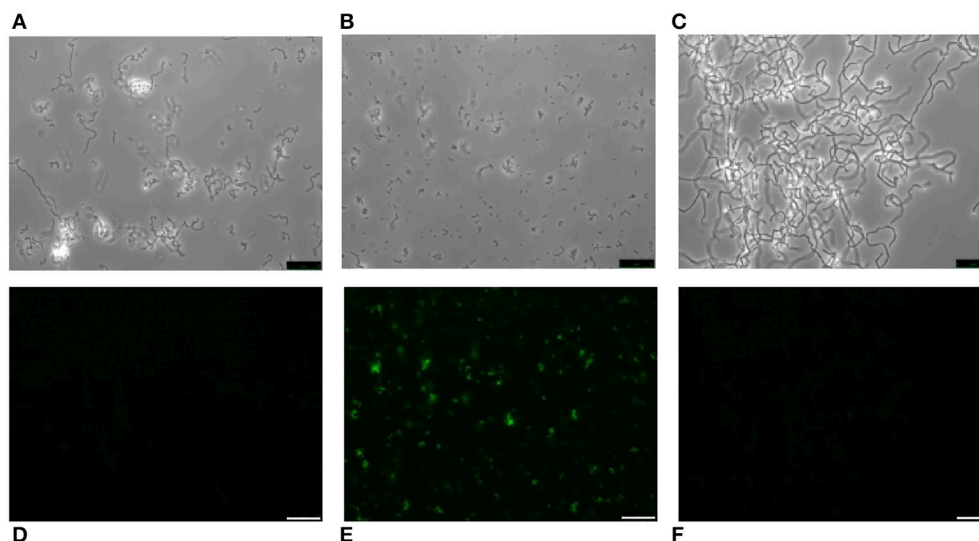
**FIGURE 6 |** HPA labeling of two choline-independent *S. pneumoniae* strains. Pneumococcal strains JY2190 (A,B) and P501 (C,D) were incubated in a chemically-defined medium (Cden) lacking any amino alcohol, and labeled with HPA (25  $\mu\text{g/ml}$ ). Bar = 25  $\mu\text{m}$ .

*S. mitis*<sup>T</sup> in a DNA–DNA hybridization assay (Kilian et al., 2008), slightly below the 70% level typically expected for two strains of the same species (Wayne et al., 1987). Whether strain SK137 represents a distinct species is, however, debatable, according to recent taxonomic proposals (Tindall et al., 2010).

The type strain of *S. oralis* (NCTC 11427)—the LTA structure of which is unknown—did not bind HPA (Table 1). However, a recent study has revealed that, in contrast to pnLTA, in which the structural element  $\alpha\text{GalNAc1} \rightarrow 3\beta\text{GalNAc1} \rightarrow$  is present (Gisch et al., 2013), only a  $\beta\text{GalNAc1} \rightarrow$  moiety is detectable in the *S. oralis* Uo5 LTA repeating unit (Gisch et al., 2015b). Assuming an identical LTA structure for *S. oralis*<sup>T</sup>, the lack of the  $\alpha\text{GalNAc1} \rightarrow$  residue at the non-reducing end would fit in with the absence of HPA labeling. It should be underlined that at least three biochemical variants of choline-containing TAs may occur in *S. oralis* and some *S. mitis* strains, according to recent results (Denapate et al., 2016).

Among the few other bacterial species tested, Lancefield group C and D streptococci were positive for HPA labeling. These results were expected in view of previous reports of *Streptococcus* belonging to groups C (Coligan et al., 1977; Sørensen and Henrichsen, 1987; Köhler and Nagai, 1989) and D (Kurl et al., 1989). From a diagnostic perspective, however, it is important to underline that non-pneumococcal, viridans (i.e.,  $\alpha$ -hemolytic) streptococci are usually considered as commensals and seldom cause CAP. It is well known that viridans streptococci produce a range of invasive disease in humans (e.g., infective endocarditis) and are also emerging as a cause of bloodstream infections, but mainly in immunocompromised patients (Doern and Burnham, 2010). *S. dysgalactiae* subsp. *equisimilis*, which belongs to the  $\beta$ -hemolytic group C and G pyogenic group of streptococci, also binds HPA and is also currently considered as an emergent





**FIGURE 7 |** Lack of HPA labeling in ethanolamine-containing streptococci. *S. pneumoniae* R6 cells were incubated for several generations in Cden medium containing ethanolamine (Cden-EA) (**A,D**). A portion of the culture then received choline chloride (5  $\mu$ g/ml) and incubation proceeded at 37°C for 3 h (**B,E**). The *S. mitis* SK598 strain was grown in THY medium (**C,F**). The three cultures were incubated with HPA (25  $\mu$ g/ml) (**D–F**). Bar = 25  $\mu$ m.

human pathogen. Nevertheless, and as mentioned above for viridans streptococci, it is a frequent cause of invasive disease only in patients having underlying conditions (Broyles et al., 2009).

*S. aureus* is an important opportunistic pathogen that persistently colonizes about 20% of the human population and is intermittently associated with the remainder. This organism is one of the most frequent and important human pathogens and is implicated in a range of infections, including superficial skin infections, abscesses, and food poisoning as well as life-threatening invasive diseases (Tong et al., 2015). Although, *S. aureus* possesses surface carbohydrates that might be recognized by HPA (Krivan et al., 1988; Payne et al., 1992), the molecular basis for HPA binding is not well understood. It has been shown, however, that certain methicillin-susceptible *S. aureus* isolates—in particular those belonging to the atypical sequence type (ST) 395 lineage (e.g., strain PS187; Winstel et al., 2013)—produce a glycerol TA modified with  $\alpha$ GalNAc (Winstel et al., 2014; Lee et al., 2015) that may be responsible for HPA binding. It should be noted, however, that the chemical structures of WTA and LTA of *S. aureus* are similar but not identical (Xia et al., 2010; Brown et al., 2013), that the ST of the type strain of *S. aureus* (ST8) is only distantly related to ST395, and that strain PS187 appears to be more closely related to several coagulase-negative staphylococcal species than to other *S. aureus* isolates (Winstel et al., 2013); the ST of *S. aureus* 15981 is unknown. It has been demonstrated that staphylococci can be accurately differentiated from streptococci in Gram-stained preparations (cocci in clusters and diplococci or short chains, respectively; Agger and Maki, 1978), and that *S. aureus* is more uncommon than the latter in CAP and acute meningitis cases. However, our results warrant additional studies to determine whether HPA

labeling is common among *S. aureus* isolates, particularly in methicillin-resistant isolates that represent a global health care problem (Tong et al., 2015).

In agreement with a previous report (Domenech and García, 2017), NT *H. influenzae* strains did not bind HPA; identical results are shown here for *P. aeruginosa* PAO1. Together with *S. pneumoniae*, both Gram-negative species are common in biofilms formed during acute otitis media and live-threatening, chronic respiratory diseases such as chronic obstructive pulmonary disease or cystic fibrosis (Blasi et al., 2016). The use of HPA as *S. pneumoniae*-specific lectin for fluorescence imaging should provide a powerful tool for future research on these and other relevant human pathogens forming multispecies biofilms.

HPA labeling combined with Gram staining and/or antigen detection may also constitute an appropriate combination for the rapid diagnosis of CAP and, perhaps, other conditions such as bacterial meningitis (McGill et al., 2016). This may facilitate the rapid implementation of an appropriate antibiotic regime, which is conditional on the age of the patient and the regional rate of decreased susceptibility of *S. pneumoniae* to  $\beta$ -lactam antibiotics (van de Beek et al., 2016). As a proof of concept, cultures of pneumococci, *H. influenzae* and GBS were mixed either with sheep blood or fetal bovine serum. Also using human whole blood, only pneumococci (either encapsulated or non-encapsulated) were clearly identified by HPA labeling when used at a relatively low bacterial concentration ( $2.5 \times 10^6$  cfu/ml). This concentration is lower than that frequently found in the blood of CAP patients (between  $7 \times 10^7$  and  $8 \times 10^8$  cfu/ml; Gadsby et al., 2016) or in the CSF of children with confirmed pneumococcal meningitis (median bacterial load  $\approx 5 \times 10^7$  DNA copies/ml) at the time of admittance (Roine et al., 2009). If required,



bacteria could be concentrated by centrifugation before or after staining.

There are a number of limitations to our study: (1) although the HPA labeling technique does not require a very specialized personnel, fluorescence microscopy may be unavailable in many laboratories of developing countries. To circumvent this problem, biotinylated HPA, and alkaline phosphatase- or ferritin-conjugated HPA could be employed. (2) Only seven different bacterial genera—including 15 streptococcal and 3 staphylococcal strains—, were tested. It should be noted, however, that the bacteria tested here include some of the microorganisms most frequently causing CAP and other severe diseases. (3) Although pneumococci and staphylococci are morphologically different, the finding that two *S. aureus* strains also bind HPA may sometimes represent a diagnostic drawback and deserves further research. It should be noted, however, that, for example, severe sepsis—a serious complication of CAP—is caused by *S. pneumoniae* about 100 times more frequently than by *S. aureus*, whereas methicillin-resistant *S. aureus* is an important cause of antimicrobial-resistant hospital-acquired infections worldwide and remains a public health priority in Europe (Montull et al., 2016). (4) This represents an *in vitro* study

and an appropriate evaluation of the benefits of HPA labeling for diagnostic purposes should be performed directly with clinical samples, e.g., sputum, bronchoalveolar fluid, blood, and/or CSF.

## AUTHOR CONTRIBUTIONS

MD and EG conceived and designed the experiments. MD performed the experiments. MD and EG analyzed the data and wrote the paper.

## ACKNOWLEDGMENTS

The authors thank M. H. Nam (University of Alabama at Birmingham, USA) for kindly providing strain MNZ67, M. Moscoso, J. Yuste, and P. García for helpful comments and for critically reading the manuscript, A. Burton for correcting the English version, and E. Cano and S. Ruiz for skillful technical assistance. This work was supported by a grant from the Ministerio de Economía y Competitividad (MINECO) (SAF2012-39444-C02-01). CIBER de Enfermedades Respiratorias (CIBERES) is an initiative of the Instituto de Salud Carlos III.

## REFERENCES

- Agger, W. A., and Maki, D. G. (1978). Efficacy of direct Gram stain in differentiating staphylococci from streptococci in blood cultures positive for gram-positive cocci. *J. Clin. Microbiol.* 7, 111–113.
- Ahmed, S., Shapiro, N. L., and Bhattacharyya, N. (2014). Incremental health care utilization and costs for acute otitis media in children. *Laryngoscope* 124, 301–305. doi: 10.1002/lary.24190
- Amann, R., and Fuchs, B. M. (2008). Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nat. Rev. Micro.* 6, 339–348. doi: 10.1038/nrmicro1888
- Balsells, E., Guillot, L., Nair, H., and Kyaw, M. H. (2017). Serotype distribution of *Streptococcus pneumoniae* causing invasive disease in children in the post-PCV era: a systematic review and meta-analysis. *PLoS ONE* 12:e0177113. doi: 10.1371/journal.pone.0177113
- Bergström, N., Jansson, P. -E., Kilian, M., and Sørensen, U. B. S. (2003). A unique variant of streptococcal group O-antigen (C-polysaccharide) that lacks phosphocholine. *Eur. J. Biochem.* 270, 2157–2162. doi: 10.1046/j.1432-1033.2003.03569.x
- Bergström, N., Jansson, P. E., Kilian, M., and Sørensen, U. B. S. (2000). Structures of two cell wall-associated polysaccharides of a *Streptococcus mitis* biovar 1 strain. A unique teichoic acid-like polysaccharide and the group O antigen which is a C-polysaccharide in common with pneumococci. *Eur. J. Biochem.* 267, 7147–7157. doi: 10.1046/j.1432-1327.2000.01821.x-i2
- Blasi, F., Page, C., Rossolini, G. M., Pallecchi, L., Matera, M. G., Rogliani, P., et al. (2016). The effect of N-acetylcysteine on biofilms: implications for the treatment of respiratory tract infections. *Respir. Med.* 117, 190–197. doi: 10.1016/j.rmed.2016.06.015
- Briese, T., and Hakenbeck, R. (1984). "Influence of lipoteichoic acid and choline on the autolytic enzyme activity of *Streptococcus pneumoniae*," in *Microbial Cell Wall Synthesis and Autolysis*, ed C. Nombela (Amsterdam: Elsevier Science Publishers), 201–206.
- Briles, E. B., and Tomasz, A. (1973). Pneumococcal Forssman antigen. A choline-containing LTA. *J. Biol. Chem.* 248, 6394–6397.
- Briles, E. B., and Tomasz, A. (1975). Physiological studies on the pneumococcal Forssman antigen: a choline-containing lipoteichoic acid. *J. Gen. Microbiol.* 86, 267–274. doi: 10.1099/00221287-86-2-267
- Brown, S., Santa Maria, J. P. Jr., and Walker, S. (2013). Wall teichoic acids of gram-positive bacteria. *Annu. Rev. Microbiol.* 67, 313–336. doi: 10.1146/annurev-micro-092412-155620
- Broyles, L. N., Van Beneden, C., Beall, B., Facklam, R., Shewmaker, P. L., Malpiedi, P., et al. (2009). Population-based study of invasive disease due to  $\beta$ -hemolytic streptococci of groups other than A and B. *Clin. Infect. Dis.* 48, 706–712. doi: 10.1086/597035
- Chochua, S., D'Acremont, V., Hanke, C., Alfa, D., Shak, J., Kilowoko, M., et al. (2016). Increased nasopharyngeal density and concurrent carriage of *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* are associated with pneumonia in febrile children. *PLoS ONE* 11:e0167725. doi: 10.1371/journal.pone.0167725
- Cilloniz, C., Martin-Loeches, I., Garcia-Vidal, C., San Jose, A., and Torres, A. (2016). Microbial etiology of pneumonia: epidemiology, diagnosis and resistance patterns. *Int. J. Mol. Sci.* 17:2120. doi: 10.3390/ijms17122120
- Clark, J. E. (2015). Determining the microbiological cause of a chest infection. *Arch. Dis. Child.* 100, 193–197. doi: 10.1136/archdischild-2013-305742
- Coligan, J. E., Fraser, B. A., and Kindt, T. J. (1977). A disaccharide hapten from streptococcal group C carbohydrate that cross-reacts with the Forssman glycolipid. *J. Immunol.* 118, 6–11.
- Cooling, L. (2015). Blood groups in infection and host susceptibility. *Clin. Microbiol. Rev.* 28, 801–870. doi: 10.1128/CMR.00109-14
- Dekker, J. P., and Lau, A. F. (2016). An update on the *Streptococcus bovis* group: classification, identification, and disease associations. *J. Clin. Microbiol.* 54, 1694–1699. doi: 10.1128/JCM.02977-15
- Denapate, D., Rieger, M., Köndgen, S., Brückner, R., Ochigava, I., Kappeler, P., et al. (2016). Highly variable *Streptococcus oralis* strains are common among viridans streptococci isolated from primates. *mSphere* 1:e00041-15. doi: 10.1128/mSphere.00041-15
- Derezinski, P., Klupczynska, A., Sawicki, W., Pałka, J. A., and Kokot, Z. J. (2017). Amino acid profiles of serum and urine in search for prostate cancer biomarkers: a pilot study. *Int. J. Med. Sci.* 14, 1–12. doi: 10.7150/ijms.15783
- Doern, C. D., and Burnham, C.-A. D. (2010). It's not easy being green: the viridans group streptococci, with a focus on pediatric clinical manifestations. *J. Clin. Microbiol.* 48, 3829–3835. doi: 10.1128/JCM.01563-10
- Domenech, M., and García, E. (2017). N-Acetyl-L-cysteine and cysteamine: new strategies against mixed biofilms of non-encapsulated *Streptococcus*

- pneumoniae and non-typeable *Haemophilus influenzae*. *Antimicrob. Agents Chemother.* 61:e01992-16. doi: 10.1128/AAC.01992-16
- Domenech, M., Damián, D., Ardanuy, C., Liñares, J., Fenoll, A., and García, E. (2015). Emerging, non-PCV13 serotypes 11A and 35B of *Streptococcus pneumoniae* show high potential for biofilm formation *in vitro*. *PLoS ONE* 10:e0125636. doi: 10.1371/journal.pone.0125636
- Domenech, M., García, E., and Moscoso, M. (2009). Versatility of the capsular genes during biofilm formation by *Streptococcus pneumoniae*. *Environ. Microbiol.* 11, 2542–2555. doi: 10.1111/j.1462-2920.2009.01979.x
- Domenech, M., García, E., and Moscoso, M. (2011). *In vitro* destruction of *Streptococcus pneumoniae* biofilms with bacterial and phage peptidoglycan hydrolases. *Antimicrob. Agents Chemother.* 55, 4144–4148. doi: 10.1128/AAC.00492-11
- Ewig, S., Birkner, N., Strauss, R., Schaefer, E., Pauletzki, J., Bischoff, H., et al. (2009). New perspectives on community-acquired pneumonia in 388 406 patients. Results from a nationwide mandatory performance measurement programme in healthcare quality. *Thorax* 64, 1062–1069. doi: 10.1136/thx.2008.109785
- Forteschi, M., Zinellu, A., Assaretti, S., Mangoni, A. A., Pintus, G., Carru, C., et al. (2016). An isotope dilution capillary electrophoresis/tandem mass spectrometry (CE-MS/MS) method for the simultaneous measurement of choline, betaine, and dimethylglycine concentrations in human plasma. *Anal. Bioanal. Chem.* 408, 7505–7512. doi: 10.1007/s00216-016-9848-6
- Frölich, L., Dirr, A., Götz, M. E., Gsell, W., Reichmann, H., Riederer, P., et al. (1998). Acetylcholine in human CSF: methodological considerations and levels in dementia of Alzheimer type. *J. Neural Transm.* 105, 961–973. doi: 10.1007/s007020050105
- Gadsby, N. J., Russell, C. D., McHugh, M. P., Mark, H., Conway Morris, A., Laurenson, I. F., et al. (2016). Comprehensive molecular testing for respiratory pathogens in community-acquired pneumonia. *Clin. Infect. Dis.* 62, 817–823. doi: 10.1093/cid/civ1214
- García-Cobos, S., Moscoso, M., Pumarola, F., Arroyo, M., Lara, N., Pérez-Vázquez, M., et al. (2014). Frequent carriage of resistance mechanisms to  $\beta$ -lactams and biofilm formation in *Haemophilus influenzae* causing treatment failure and recurrent otitis media in young children. *J. Antimicrob. Chemother.* 69, 2394–2399. doi: 10.1093/jac/dku158
- Geno, K. A., Saad, J. S., and Nahm, M. H. (2017). Discovery of novel pneumococcal serotype 35D, a natural WciG-deficient variant of serotype 35B. *J. Clin. Microbiol.* 55, 1416–1425. doi: 10.1128/JCM.00054-17
- Gisch, N., Kohler, T., Ulmer, A. J., Müthing, J., Pribyl, T., Fischer, K., et al. (2013). Structural reevaluation of *Streptococcus pneumoniae* lipoteichoic acid and new insights into its immunostimulatory potency. *J. Biol. Chem.* 288, 15654–15667. doi: 10.1074/jbc.M112.446963
- Gisch, N., Peters, K., Zähringer, U., and Vollmer, W. (2015a). “The pneumococcal cell wall,” in *Streptococcus pneumoniae: Molecular Mechanisms of Host-Pathogen Interactions*, eds J. Brown, S. Hammerschmidt, and C. Orihuela (San Diego, CA: Elsevier, Inc.), 145–167.
- Gisch, N., Schwudke, D., Thomsen, S., Heb, N., Hakenbeck, R., and Denapate, D. (2015b). Lipoteichoic acid of *Streptococcus oralis* Uo5: a novel biochemical structure comprising an unusual phosphorylcholine substitution pattern compared to *Streptococcus pneumoniae*. *Sci. Rep.* 5:16718. doi: 10.1038/srep16718
- Giudicelli, S., and Tomasz, A. (1984). “Inhibition of the *in vitro* and *in vivo* activity of the pneumococcal autolytic enzyme by choline and phosphorylcholine,” in *Microbial Cell Wall Synthesis and Autolysis*, ed C. Nombela (Amsterdam: Elsevier Science Publishers), 207–212.
- González, A., Llull, D., Morales, M., García, P., and García, E. (2008). Mutations in the *tacF* gene of clinical strains and laboratory transformants of *Streptococcus pneumoniae*: impact on choline auxotrophy and growth rate. *J. Bacteriol.* 190, 4129–4138. doi: 10.1128/JB.01991-07
- Hajjeh, R., Mulholland, K., Schuchat, A., and Santosham, M. (2013). Progress towards demonstrating the impact of *Haemophilus influenzae* type b conjugate vaccines globally. *J. Pediatr.* 163, S1–S3. doi: 10.1016/j.jpeds.2013.03.022
- Haq, I. J., Battersby, A. C., Eastham, K., and McKean, M. (2017). Community acquired pneumonia in children. *BMJ* 356:j686. doi: 10.1136/bmj.j686
- Hoskins, J., Alborn, W. E. Jr., Arnold, J., Blaszcak, L. C., Burgett, S., DeHoff, B. S., et al. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* 183, 5709–5717. doi: 10.1128/JB.183.19.5709-5717.2001
- Jenkin, C. R. (1963). Heterophile antigens and their significance in the host-parasite relationship. *Adv. Immunol.* 3, 351–376. doi: 10.1016/S0065-2776(08)60816-0
- Keller, L. E., Robinson, D. A., and McDaniel, L. S. (2016). Nonencapsulated *Streptococcus pneumoniae*: emergence and pathogenesis. *mBio* 7:e01792-15. doi: 10.1128/mBio.01792-15
- Kilian, M., Poulsen, K., Blomqvist, T., Håvarstein, L. S., Bek-Thomsen, M., Tettelin, H., et al. (2008). Evolution of *Streptococcus pneumoniae* and its close commensal relatives. *PLoS ONE* 3:e2683. doi: 10.1371/journal.pone.0002683
- Kilpper-Bälz, R., Wenzig, P., and Schleifer, K. H. (1985). Molecular relationships and classification of some viridans streptococci as *Streptococcus oralis* and emended description of *Streptococcus oralis* (Bridge and Sneath 1982). *Int. J. Syst. Bacteriol.* 35, 482–488. doi: 10.1099/00207713-35-4-482
- Kjos, M., Aprianto, R., Fernandes, V. E., Andrew, P. W., van Strijp, J. A. G., Nijland, R., et al. (2015). Bright fluorescent *Streptococcus pneumoniae* for live-cell imaging of host-pathogen interactions. *J. Bacteriol.* 197, 807–818. doi: 10.1128/JB.02221-14
- Kobayashi, Y., Tateno, H., Ogawa, H., Yamamoto, K., and Hirabayashi, J. (2014). Comprehensive list of lectins: origins, natures, and carbohydrate specificities. *Methods Mol. Biol.* 1200, 555–577. doi: 10.1007/978-1-4939-1292-6\_45
- Köhler, W., and Nagai, T. (1989). Reactions of the lectin anti-A<sub>HP</sub> from the edible snail *Helix pomatia* with N-acetyl-D-galactosamine of streptococci. *Kitasato Arch. Exp. Med.* 62, 107–113.
- Krivan, H. C., Roberts, D. D., and Ginsburg, V. (1988). Many pulmonary pathogenic bacteria bind specifically to the carbohydrate sequence GalNAc $\beta$ 1-4Gal found in some glycolipids. *Proc. Natl. Acad. Sci. U.S.A.* 85, 6157–6161. doi: 10.1073/pnas.85.16.6157
- Kruse, T., Reiber, H., and Neuhoof, V. (1985). Amino acid transport across the human blood-CSF barrier. An evaluation graph for amino acid concentrations in cerebrospinal fluid. *J. Neurol. Sci.* 70, 129–138. doi: 10.1016/0022-510X(85)90082-6
- Kurl, D. N., Haataja, S., and Finne, J. (1989). Hemagglutination activities of group B, C, D, and G streptococci: demonstration of novel sugar-specific cell-binding activities in *Streptococcus suis*. *Infect. Immun.* 57, 384–389.
- Lacks, S., and Hotchkiss, R. D. (1960). A study of the genetic material determining an enzyme activity in *Pneumococcus*. *Biochim. Biophys. Acta* 39, 508–518. doi: 10.1016/0006-3002(60)90205-5
- Lanie, J. A., Ng, W. -L., Kazmierczak, K. M., Andrzejewski, T. M., Davidsen, T. M., Wayne, K. J., et al. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J. Bacteriol.* 189, 38–51. doi: 10.1128/JB.01148-06
- Lee, J. -H., Kim, N. -H., Winstel, V., Kurokawa, K., Larsen, J., An, J. -H., et al. (2015). Surface glycopolymers are crucial for *in vitro* anti-wall teichoic acid IgG-mediated complement activation and opsonophagocytosis of *Staphylococcus aureus*. *Infect. Immun.* 83, 4247–4255. doi: 10.1128/IAI.00767-15
- Liu, W., Röder, H. L., Madsen, J. S., Bjarnsholt, T., Sørensen, S. J., and Burmølle, M. (2016). Interspecific bacterial interactions are reflected in multispecies biofilm spatial organization. *Front. Microbiol.* 7:1366. doi: 10.3389/fmicb.2016.01366
- López, R., and García, E. (2004). Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiol. Rev.* 28, 553–580. doi: 10.1016/j.femsre.2004.05.002
- Matsui, T., Hamako, J., Ozeki, Y., and Titani, K. (2001). Comparative study of blood group-recognizing lectins toward ABO blood group antigens on neoglycoproteins, glycoproteins and complex-type oligosaccharides. *Biochim. Biophys. Acta* 1525, 50–57. doi: 10.1016/S0304-4165(00)00170-7
- McGill, F., Heyderman, R. S., Panagiotou, S., Tunkel, A. R., and Solomon, T. (2016). Acute bacterial meningitis in adults. *Lancet* 388, 3036–3047. doi: 10.1016/S0140-6736(16)30654-7
- Momeni, B., Brileya, K. A., Fields, M. W., and Shou, W. (2013). Strong inter-population cooperation leads to partner intermixing in microbial communities. *eLife* 2:e00230. doi: 10.7554/eLife.00230
- Monasta, L., Ronfani, L., Marchetti, F., Montico, M., Vecchi Brumatti, L., Bavcar, A., et al. (2012). Burden of disease caused by otitis media: systematic review and global estimates. *PLoS ONE* 7:e36226. doi: 10.1371/journal.pone.0036226

- Montull, B., Menéndez, R., Torres, A., Reyes, S., Méndez, R., Zalacáin, R., et al. (2016). Predictors of severe sepsis among patients hospitalized for community-acquired pneumonia. *PLoS ONE* 11:e0145929. doi: 10.1371/journal.pone.0145929
- Moscoso, M., Domenech, M., and García, E. (2010). Vancomycin tolerance in clinical and laboratory *Streptococcus pneumoniae* isolates depends on reduced enzyme activity of the major LytA autolysin or cooperation between CiaH histidine kinase and capsular polysaccharide. *Mol. Microbiol.* 77, 1052–1064. doi: 10.1111/j.1365-2958.2010.07271.x
- Ngo, C. C., Massa, H. M., Thornton, R. B., and Cripps, A. W. (2016). Predominant bacteria detected from the middle ear fluid of children experiencing otitis media: a systematic review. *PLoS ONE* 11:e0150949. doi: 10.1371/journal.pone.0150949
- Ogawa, S., Hattori, K., Sasayama, D., Yokota, Y., Matsumura, R., Matsuo, J., et al. (2015). Reduced cerebrospinal fluid ethanolamine concentration in major depressive disorder. *Sci. Rep.* 5:7796. doi: 10.1038/srep07796
- Pai, S., Enoch, D. A., and Aliyu, S. H. (2015). Bacteremia in children: epidemiology, clinical diagnosis and antibiotic treatment. *Expert Rev. Anti Infect. Ther.* 13, 1073–1088. doi: 10.1586/14787210.2015.1063418
- Park, I. H., Kim, K.-H., Andrade, A. L., Briles, D. E., McDaniel, L. S., and Nahm, M. H. (2012). Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene *pspK*. *mBio* 3:e00035-12. doi: 10.1128/mBio.00035-12
- Payne, M. J., Campbell, S., Patchett, R. A., and Kroll, R. G. (1992). The use of immobilized lectins in the separation of *Staphylococcus aureus*, *Escherichia coli*, *Listeria* and *Salmonella* spp. from pure cultures and foods. *J. Appl. Bacteriol.* 73, 41–52. doi: 10.1111/j.1365-2672.1992.tb04967.x
- Rane, L., and Subbarow, Y. (1940). Nutritional requirements of the pneumococcus. I. Growth factors for types I, II, V, VII, VIII. *J. Bacteriol.* 40, 695–704.
- Røder, H. L., Sørensen, S. J., and Burmølle, M. (2016). Studying bacterial multispecies biofilms: where to start? *Trends Microbiol.* 24, 503–513. doi: 10.1016/j.tim.2016.02.019
- Roine, I., Saukkoripi, A., Leinonen, M., Peltola, H., and LatAm Meningitis Study Group (2009). Microbial genome count in cerebrospinal fluid compared with clinical characteristics in pneumococcal and *Haemophilus influenzae* type b meningitis in children. *Diagn. Microbiol. Infect. Dis.* 63, 16–23. doi: 10.1016/j.diagmicrobio.2008.09.005
- Schimak, M. P., Kleiner, M., Wetzel, S., Liebeke, M., Dubilier, N., and Fuchs, B. M. (2016). MiL-FISH: multilabeled oligonucleotides for fluorescence *in situ* hybridization improve visualization of bacterial cells. *Appl. Environ. Microbiol.* 82, 62–70. doi: 10.1128/AEM.02776-15
- Seo, H. S., Cartee, R. T., Pritchard, D. G., and Nahm, M. H. (2008). A new model of pneumococcal lipoteichoic acid structure resolves biochemical, biosynthetic, and serologic inconsistencies of the current model. *J. Bacteriol.* 190, 2379–2387. doi: 10.1128/JB.01795-07
- Shiroma, A., Terabayashi, Y., Nakano, K., Shimoji, M., Tamotsu, H., Ashimine, N., et al. (2015). First complete genome sequences of *Staphylococcus aureus* subsp. *aureus* Rosenbach 1884 (DSM 20231T), determined by PacBio single-molecule real-time technology. *Genome Announc.* 3:e00800-15. doi: 10.1128/genomeA.00800-15
- Siddiqui, B., and Hakomori, S.-I. (1971). A revised structure for the Forssman glycolipid hapten. *J. Biol. Chem.* 246, 5766–5769.
- Sørensen, U. B., and Henriksen, J. (1987). Cross-reactions between pneumococci and other streptococci due to C polysaccharide and F antigen. *J. Clin. Microbiol.* 25, 1854–1859.
- Tindall, B. J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0
- Tomasz, A. (1968). Biological consequences of the replacement of choline by ethanolamine in the cell wall of *Pneumococcus*: chain formation, loss of transformability, and loss of autolysis. *Proc. Natl. Acad. Sci. U.S.A.* 59, 86–93. doi: 10.1073/pnas.59.1.86
- Tomasz, A., Westphal, M., Briles, E. B., and Fletcher, P. (1975). On the physiological functions of teichoic acids. *J. Supramol. Struct.* 3, 1–16. doi: 10.1002/jss.400030102
- Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L., and Fowler, V. G. Jr. (2015). *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* 28, 603–661. doi: 10.1128/CMR.00134-14
- Torres, A., Lee, N., Cilloniz, C., Vila, J., and Van der Eerden, M. (2016). Laboratory diagnosis of pneumonia in the molecular age. *Eur. Respir. J.* 48, 1764–1778. doi: 10.1183/13993003.01144-2016
- Valle, J., Toledo-Arana, A., Berasain, C., Ghigo, J.-M., Amorena, B., Penadés, J. R., et al. (2003). SarA and not  $\sigma^B$  is essential for biofilm development by *Staphylococcus aureus*. *Mol. Microbiol.* 48, 1075–1087. doi: 10.1046/j.1365-2958.2003.03493.x
- van de Beek, D., Cabellos, C., Dzupova, O., Esposito, S., Klein, M., Kloek, A. T., et al. (2016). ESCMID guideline: diagnosis and treatment of acute bacterial meningitis. *Clin. Microbiol. Infect.* 22, S37–S62. doi: 10.1016/j.cmi.2016.01.007
- Vila, J., Gómez, M. D., Salavert, M., and Bosch, J. (2017). Methods of rapid diagnosis in clinical microbiology: clinical needs. *Enferm. Infecc. Microbiol. Clin.* 35, 41–46. doi: 10.1016/j.eimc.2016.11.004
- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Evol. Microbiol.* 37, 463–464. doi: 10.1099/00207713-37-4-463
- Winstel, V., Liang, C., Sanchez-Carballo, P., Steglich, M., Munar, M., Bröker, B. M., et al. (2013). Wall teichoic acid structure governs horizontal gene transfer between major bacterial pathogens. *Nat. Commun.* 4:2345. doi: 10.1038/ncomms3345
- Winstel, V., Xia, G., and Peschel, A. (2014). Pathways and roles of wall teichoic acid glycosylation in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 304, 215–221. doi: 10.1016/j.ijmm.2013.10.009
- World Health Organization (2016). *Pneumonia*. Geneva: World Health Organization. Available online at: <http://www.who.int/mediacentre/factsheets/fs331/en/>
- Wu, A. M., and Sugii, S. (1991). Coding and classification of D-galactose, N-acetyl-D-galactosamine, and  $\beta$ -D-Galp-[1 $\rightarrow$ 3(4)]- $\beta$ -D-GlcpNAc, specificities of applied lectins. *Carbohydr. Res.* 213, 127–143. doi: 10.1016/S0008-6215(00)90604-9
- Xia, G., Kohler, T., and Peschel, A. (2010). The wall teichoic acid and lipoteichoic acid polymers of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 300, 148–154. doi: 10.1016/j.ijmm.2009.10.001
- Yamamoto, M., Cid, E., and Yamamoto, F. (2012). Molecular genetic basis of the human Forssman glycolipid antigen negativity. *Sci. Rep.* 2:975. doi: 10.1038/srep00975
- Yother, J., Leopold, K., White, J., and Fischer, W. (1998). Generation and properties of a *Streptococcus pneumoniae* mutant which does not require choline or analogs for growth. *J. Bacteriol.* 180, 2093–2101

**Conflict of Interest Statement:** The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer ME declared a shared affiliation, though no other collaboration, with one of the authors MD to the handling Editor, who ensured that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Domenech and García. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Molecular Mechanisms That Contribute to Horizontal Transfer of Plasmids by the Bacteriophage SPP1

Ana Valero-Rello<sup>1,2†</sup>, María López-Sanz<sup>1†</sup>, Alvaro Quevedo-Olmos<sup>1</sup>, Alexei Sorokin<sup>2</sup> and Silvia Ayora<sup>1\*</sup>

<sup>1</sup> Department of Microbial Biotechnology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
United States

### Reviewed by:

Maite Muniesa,  
University of Barcelona, Spain  
Elisabeth Grohmann,  
Beuth University of Applied Sciences,  
Germany

### \*Correspondence:

Silvia Ayora  
sayora@cnb.csic.es

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 18 July 2017

**Accepted:** 06 September 2017

**Published:** 22 September 2017

### Citation:

Valero-Rello A, López-Sanz M,  
Quevedo-Olmos A, Sorokin A and  
Ayora S (2017) Molecular  
Mechanisms That Contribute  
to Horizontal Transfer of Plasmids by  
the Bacteriophage SPP1.  
Front. Microbiol. 8:1816.  
doi: 10.3389/fmicb.2017.01816

Natural transformation and viral-mediated transduction are the main avenues of horizontal gene transfer in Firmicutes. *Bacillus subtilis* SPP1 is a generalized transducing bacteriophage. Using this lytic phage as a model, we have analyzed how viral replication and recombination systems contribute to the transfer of plasmid-borne antibiotic resistances. Phage SPP1 DNA replication relies on essential phage-encoded replisome organizer (G38P), helicase loader (G39P), hexameric replicative helicase (G40P), recombinase (G35P) and in less extent on the partially dispensable 5'→3' exonuclease (G34.1P), the single-stranded DNA binding protein (G36P) and the Holliday junction resolvase (G44P). Correspondingly, the accumulation of linear concatemeric plasmid DNA, and the formation of transducing particles were blocked in the absence of G35P, G38P, G39P, and G40P, greatly reduced in the G34.1P, G36P mutants, and slightly reduced in G44P mutants. In contrast, establishment of injected linear plasmid DNA in the recipient host was independent of viral-encoded functions. DNA homology between SPP1 and the plasmid, rather than a viral packaging signal, enhanced the accumulation of packagable plasmid DNA. The transfer efficiency was also dependent on plasmid copy number, and rolling-circle plasmids were encapsidated at higher frequencies than theta-type replicating plasmids.

**Keywords:** horizontal gene transfer, plasmid transduction, SPP1, bacteriophages, antibiotic resistance

## INTRODUCTION

Bacteriophage-mediated horizontal gene transfer enhances bacterial adaptive responses to environmental changes, and it is one of the mechanisms responsible for the rapid spread of antibiotic resistance, bacterial virulence and pathogenicity (Canchaya et al., 2003; Brussow et al., 2004; Brown-Jaque et al., 2015; Penades et al., 2015; Touchon et al., 2017). Bacteriophages, or simply phages, play active roles in the specialized mobilization of discrete chromosomal regions (specialized transduction), and also with significant efficiency can transfer any chromosomal segment or plasmid DNA (generalized transduction). The difference between these two transduction modes is that specialized transduction is the consequence of the faulty excision of the prophage from the bacterial chromosome, resulting into packaging of phage DNA as well as

**Abbreviations:** PFGE, pulsed field gel electrophoresis; RCR, rolling circle replication; SPP1, *B. subtilis* bacteriophage SPP1; *sus*, suppressor sensitive (mutation); TR, theta replication; *ts*, thermosensitive; *wt*, wild type.



adjacent DNA from the bacterial chromosome (Canchaya et al., 2003; Brussow et al., 2004; Penades et al., 2015; Touchon et al., 2017). In generalized transduction, phage DNA mispackaging occurs, and the viral packaging machinery uses chromosomal or plasmid DNA as a substrate for DNA packaging into the empty proheads instead of viral DNA (Ikeda and Tomizawa, 1965; Viret et al., 1991). Generalized transduction, which is recognized as a widespread mechanism for the transfer of any gene from one bacterium to another, was originally reported in  $\gamma$ -proteobacteria (Zinder and Lederberg, 1952; Lennox, 1955), and it has been also reported in many Gram-positive pathogens (Maslanova et al., 2013; Giovanetti et al., 2014; Winstel et al., 2015). The majority of generalized transducing phages package their DNA by the headful packaging mechanism (*pac* phages). One remarkable event related to this, is the encapsidation of pathogenicity islands, as it occurs with the *Staphylococcus aureus* pathogenicity islands (SaPIs). SaPIs have developed elegant strategies to hijack the phage machinery to use it for their own transfer (Penades et al., 2015). Most SaPI helper phages identified to date are *pac* phages, and many well-studied SaPIs are packaged by the headful mechanism (Ruzin et al., 2001). Despite its importance in spreading antibiotic resistances and virulence, the mechanisms that occur inside the cell and lead to the erroneous encapsidation of foreign DNA upon phage infection remain largely unexplored.

SPP1 is a 44-kb virulent *Bacillus subtilis* phage that can carry out generalized transduction (plasmid and chromosomal) with a significant frequency (Yasbin and Young, 1974; Ferrari et al., 1978; Canosi et al., 1982). The SPP1 replication and packaging machineries have been studied in deep (Alonso et al., 2006; Lo Piano et al., 2011; Oliveira et al., 2013). SPP1 DNA replication starts by the theta mode when the replisome organizer, G38P, binds to the replication origin, *oriL* (Pedre et al., 1994; Missich et al., 1997; Seco and Ayora, 2017). Then, the phage helicase loader (G39P) recruits the replicative hexameric helicase (G40P). The viral helicase recruits the host-encoded primase (DnaG) and DnaX, which is a subunit of the clamp loader (Pedre et al., 1994; Ayora et al., 1999; Martinez-Jimenez et al., 2002), so that a full replisome is loaded at the phage origin. SPP1 replication uses the host replicase holoenzyme and topoisomerases from the host (Seco et al., 2013; Seco and Ayora, 2017). After one or two rounds of theta-type replication (TR), it shifts to concatemeric (sigma-type) DNA replication in a process driven by recombination (Lo Piano et al., 2011). Two viral proteins may participate in this shift, the ATP-independent single-strand annealing recombinase (G35P) and its partner, the 5'→3' exonuclease (G34.1P) (Ayora et al., 2002; Martinez-Jimenez et al., 2005). In the shift to concatemeric DNA replication, G38P, bound to *oriR*, or working as a pre-primosome organizer (like the bacterial PriA enzyme), may restart DNA replication at stalled or paused replication forks (Seco et al., 2013). SPP1 codes for two other proteins involved in DNA replication and recombination: the G36P and G44P proteins. G36P is a single-stranded DNA binding protein (SSB), and G44P is a Holliday junction resolvase of the RusA family, which recognizes and cleaves a variety of recombination intermediates (Martinez-Jimenez et al., 2005; Zecchi et al., 2012). Biochemical assays showed that G36P is crucial for SPP1 DNA replication *in vitro*, but it can be substituted

by host-encoded SSB (known as SsbA) (Seco et al., 2013). The role of G44P in SPP1 replication is thought to be the processing of the stalled replication fork, which may trigger the shift to the sigma-type or concatemeric DNA replication. This type of DNA replication is essential to generate the concatemeric DNA, which is the substrate for encapsidation. Viral replication and packaging are sequential and in some way coupled events. SPP1 encapsidates linear double-strand (ds) DNA into an empty prohead by a processive (~4 sequential packaging cycles) headful packaging mechanism, using the linear head-to-tail concatemer as a substrate (Oliveira et al., 2013). This is consistent with the observation that an *in vitro* DNA packaging system efficiently packaged mature SPP1 DNA as well as linear plasmid DNA, but no DNA packaging could be detected when circular DNA was the substrate for encapsidation (Oliveira et al., 2005). SPP1 packaging is initiated with the recognition of the specific *pac* region by the terminase small subunit, G1P, and the sequence specific cleavage at the *pac* sequence (CTATTGCGG↓C) by the terminase large subunit, G2P (Chai et al., 1992, 1995, 1997). This generates the first DNA end to be encapsidated (Chai et al., 1992; Gual et al., 2000; Camacho et al., 2003). A sequence independent cleavage, at 104% of the genome (headful cleavage), terminates one packaging round, generating a new starting point for another one (Chai et al., 1995; Camacho et al., 2003). Hence, the first cleavage in the concatemeric SPP1 DNA occurs specifically at *pac*, whereas the next ones do not (Gual et al., 2000).

In addition to package viral DNA, SPP1 is able to encapsidate chromosomal or plasmid DNA. However, some differences were observed with these two substrates. Rolling-circle replicating plasmids could be transduced at a frequency much higher than chromosomal DNA (Ferrari et al., 1978; Deichelbohrer et al., 1985), and an explanation for this could be that the copy number of plasmids in the cell is higher than that of the chromosome. Alternatively, another possibility could be that the replication mode influences the transduction frequency. It was also observed that the frequency of transduction of pUB110 and pC194 naturally occurring plasmids was enhanced 100- to 1000-fold by the presence of inserts homologous to the transducing phage DNA (Deichelbohrer et al., 1985). This homology-facilitated plasmid transduction was independent of the host RecA (Canosi et al., 1982; Deichelbohrer et al., 1985). In contrast, another report showed that SPP1 mediated chromosomal transduction was reduced 30-fold in cells having mutations in host functions involved in homologous recombination, such as RecA, RecU, and RecF (Ferrari et al., 1978). These differences, which were observed between plasmid and chromosomal transduction in the SPP1 system motivated us to analyze in deep and throughout the manuscript the influence of the replication mode and of the plasmid copy number in plasmid generalized transduction. In addition, we have analyzed the role of phage recombination and replication proteins. We show that in absence of G35P, G38P, G39P, or G40P linear plasmid transduction is blocked. In contrast, establishment of injected linear plasmid DNA in the recipient host was independent of viral-encoded functions. The transfer efficiency was found to be dependent on homology to phage DNA, plasmid copy number, and replication mechanism.

## MATERIALS AND METHODS

### Bacterial Strains and Plasmids

*Bacillus subtilis* BG214 (*trpCE metA5 amyE1 ytsJ1 rsbV37 xre1 xkdA1 att<sup>SPB</sup> att<sup>ICEBs1</sup>*) and its isogenic derivative BG295 (*sup3*) were used. They lack the *ICEBs1* integrative conjugative element as well as prophage PBSX, and PBSX prophage cannot be induced (Kidane et al., 2009). The plasmids used are derivatives of pHP13, pUB110, pBT233 or pNDH33 (Table 1). To construct pBT233N, the pUB110 neomycin resistance gene was cloned into *Ava*I-linearized pBT233. Different regions of the SPP1 genome were cloned into the *Hpa*I site of the pBT233N plasmid as indicated in Table 1. pHP13 derivatives were kindly provided by J. C. Alonso (CNB-CSIC). Plasmid pBT400 is a pHP13 derivative bearing an *Eco*RI-*Sal*I fragment of SPP1 DNA. Different SPP1 DNA fragments were cloned into *Xba*I- or *Sma*I-cleaved pNDH33 DNA, rendering pNDH33-1300 and pNDH33-*pac* (Table 1).

### SPP1 Phages

The SPP1 phages used in this work are listed in Table 2, including those (*sus19*, *sus53*, *sus109*, tsB3, and SPP1ΔA) previously described (Chai et al., 1992; Pedre et al., 1994; Zecchi et al., 2012).

The SPP1 tsI20F mutant was sequenced and it was found that the mutation that conferred thermosensitivity (ts), P159S, mapped in gene 35, rather than in gene 34.1, as it was previously suggested after genetic mapping (Burger and Trautner, 1978). This phage was used to construct the SPP1 *sus35* mutant. First, a lysine codon (the 10th codon in the gene 35) was replaced by an ochre (UAA) stop codon by site-directed mutagenesis using plasmid pCB610 as template (a pHP13 derivative containing SPP1 genes 34.4 to 35) and the Quickchange protocol. After sequencing confirmation the resulting plasmid (pHP13-G35P-ochre) was introduced into BG295 cells by transformation. BG295 cells bearing pHP13-G35P-ochre plasmid were infected with SPP1 tsI20F phage at 30°C for 2 h. The resulting phage lysate was used to infect BG295 cells at non-permissive temperature to

**TABLE 1** | Plasmids used in this work.

Plasmids	Plasmid characteristics	Reference
pC194	Natural rolling circle replicating (RCR) plasmid, 2.9-kb	Horinouchi and Weisblum, 1982; Alonso and Trautner, 1985
pHP13	RCR plasmid derivative of pTA1060, 4.9-kb	Haima et al., 1987
pBT163 (pHP13- <i>pac</i> )	pHP13 derivative containing SPP1 DNA including <i>pac</i> (2675 bp cloned, coordinates 43778–44010 and 1–2439)	Chai et al., 1992
pBT271 (pHP13- <i>oriL</i> )	pHP13 derivative containing SPP1 DNA including <i>oriL</i> (2975 bp, coordinates 33875–36850)	Chai et al., 1993
pBT400 (pHP13-800)	pHP13 derivative containing SPP1 DNA (864 bp, coordinates 3225–4089)	This work
pUB110	Natural RCR plasmid, 4.5-kb	Leonhardt, 1990
pUB110-cop1	pUB110 derivative, lower copy number	Leonhardt, 1990
pBG55 (pUB110-3600)	pUB110 derivative containing SPP1 DNA (3639 bp, coordinates 23117–26756)	Deichelbohrer et al., 1985
pBT233	Theta replicating (TR) plasmid, 9-kb	Ceglowski et al., 1993a
pBT233N	pBT233 derivative containing the 1304 bp neomycin resistance gene (N) from pUB110	This work
pBT233N-400	pBT233N derivative containing SPP1 DNA (414 bp, coordinates 32562–32976)	This work
pBT233N-1300	pBT233N derivative containing SPP1 DNA (1340 bp, coordinates 25051–26391)	This work
pBT233N- <i>oriL</i>	pBT233N derivative containing SPP1 <i>oriL</i> DNA (350 bp, coordinates 35801–36151)	This work
pBT233N- <i>pac</i>	pBT233N derivative containing SPP1 <i>pac</i> DNA (412 bp, coordinates 43689–44010 and 1–70)	This work
pNDH33	TR plasmid derivative of pBS72, 8.1-kb	Titok et al., 2003
pNDH33-1300	pNDH33 derivative containing SPP1 DNA (1340 bp, coordinates 25051–26391)	This work
pNDH33- <i>pac</i>	pNDH33 derivative containing SPP1 <i>pac</i> DNA (412 bp, coordinates 43689–44010 and 1–70)	This work

**TABLE 2** | SPP1 phages used in this work.

Genotype	Name	Activity / type of mutant	Reference
wt	SPP1wt	Wild type	
34.1 <sup>−</sup>	<i>sus34.1</i>	Exonuclease, ochre mutant (OM)	This work
35 <sup>−</sup>	<i>sus35</i>	Recombinase, OM	This work
35 <sup>−</sup>	tsI20F	Recombinase/thermosensitive (ts) mutant	This work
36 <sup>−</sup>	<i>sus36</i>	ssDNA binding protein, OM	This work
38 <sup>−</sup>	tsB3	Replisome organizer/ts mutant	Pedre et al., 1994
39 <sup>−</sup>	<i>sus53</i>	Helicase loader, OM	Pedre et al., 1994
40 <sup>−</sup>	<i>sus109</i>	Helicase, OM	Pedre et al., 1994
44 <sup>−</sup>	SPP1ΔA	Deletion mutant lacking Holliday junction resolvase	Zecchi et al., 2012
2 <sup>−</sup>	<i>sus19</i>	Terminase large subunit, OM	Chai et al., 1992

obtain the recombinant phages. They were picked from Luria-Bertani (LB) plates supplemented with 10 mM MgCl<sub>2</sub> (LB-Mg<sup>+</sup>) incubated at 50°C. The amplified phage was sequenced to confirm that phages had acquired the ochre mutation in gene 35, and that it had reverted to wt the tsI20F mutation. The resulting mutant phage, containing the ochre codon, was named SPP1 *sus35*.

The 37th codon (Lys) in gene 36 was replaced by an ochre (UAA) stop codon in a pHP13 derivative containing SPP1 genes 34.4 to 37. The SPP1 *sus34.1* mutant was generated by replacing, in a pHP13 derivative containing SPP1 genes 34.1 to 35, the 31th codon (AAA) of gene 34.1 by an ochre (UAA) stop codon. The SPP1 *sus36* and *sus34.1* mutants were then generated by homologous recombination between the SPP1 tsI20F phage and these plasmids carrying the stop ochre codon into the gene to be mutated, as described above. The accuracy of the resulting mutant phages was confirmed by sequencing.

SPP1wt, SPP1ΔA phages and the thermosensitive phages (tsI20F, and tsB3) were amplified in BG214 cells grown at 37°C or 30°C in LB-Mg<sup>+</sup>, whereas the *sus* phages were routinely amplified in the suppressor strain BG295 (*sup3*) at 37°C.

## Preparation of Transducing Lysates

Transducing lysates were obtained by infecting with the different SPP1 phages, at a multiplicity of infection (MOI) of 10, *B. subtilis* BG214 cells bearing the indicated plasmids, grown up to mid-exponential phase in LB-Mg<sup>+</sup> and appropriated antibiotics. Aliquots were taken at different post-infection times for DNA analysis and processed as described below. The cultures were centrifuged after 90 min of infection (14,000 rpm, 5 min), and the supernatants were filtered through 0.45 μm filters to remove donor cells. Under these growth conditions *B. subtilis* cells are not competent, so that DNase I treatment was not required. Phage lysates were titrated on BG214 cells or BG295 cells before use and were stored at 4°C.

## Plasmid Transduction

Exponentially growing recipient *B. subtilis* BG214 or BG295 cells (OD<sub>560</sub> = 0.4) grown at 37°C in LB-Mg<sup>+</sup>, were infected with the transducing phage lysate at MOI of 1. Phages were allowed to be absorbed for 5 min, and then the non-absorbed phages were removed by centrifugation. Cell pellets were washed and finally resuspended in 1 ml LB. Appropriate dilutions were plated in selective LB-agar plates containing the respective antibiotics, and incubated overnight at 37°C to quantify the number of transductants. As a control, 1 ml of the recipient host was plated to discard the appearance of spontaneous resistant colonies. In another LB-agar plate with antibiotic the same amount of the stock transducing lysate was plated without recipient cells, to discard a contamination with donor cells.

## Analysis of Plasmid DNA Forms

*B. subtilis* BG214 cells bearing the different plasmids were grown at 37°C to an OD<sub>560</sub> of 0.40 in LB-Mg<sup>+</sup> media supplemented with appropriate antibiotics, and infected with a MOI of 10. Phage addition marked the time zero of our experiments. At given times, aliquots of 1ml were collected, rapidly placed in

a water-ice mixture and centrifuged for 5 min at 14,000 rpm and 4°C. The pellets were stored at −80°C. In experiments with thermosensitive phage mutants, the strains bearing plasmids were first grown at 30°C to an OD<sub>560</sub> of 0.2, transferred to 50°C and then further grown to OD<sub>560</sub> of 0.4. They were infected at 50°C, and the samples were processed as described above. Total DNA was isolated following a protocol described earlier (Viret and Alonso, 1987) with some minor modifications. Samples were resuspended in 200 μl of lysis buffer (25 mM Tris-HCl pH 8.0, 50mM glucose, 10 mM EDTA, 0.5 mg/ml lysozyme and 0.1 mg/ml RNase A). After 30 min of incubation at 30°C, Proteinase K (0.5 mg/ml) and SDS (0.8%) were added, and the mixture was further incubated for 30 min at 37°C. The lysate obtained was then treated twice with phenol and dialyzed against 20 mM Tris-HCl pH 8.0, 1 mM EDTA.

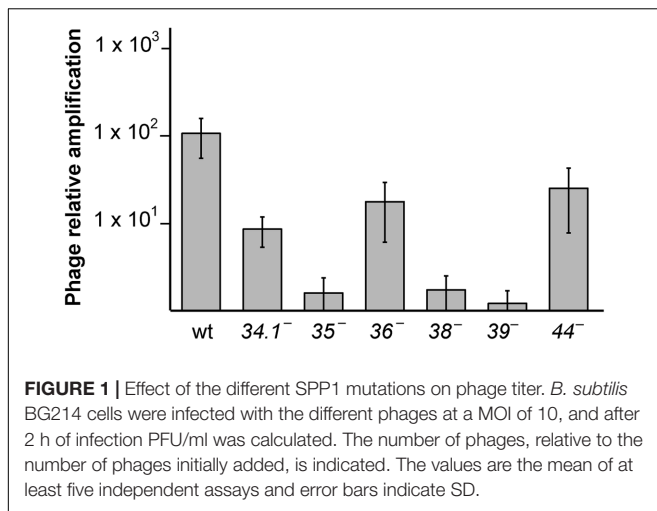
Pulsed field gel electrophoresis (PFGE) was performed on a Bio-Rad CHEF-DR II apparatus. 15 μl of samples were loaded on the 1% agarose gel. Running conditions were 5 V/cm, 0.5% TBE, 0.5–10 switch time for 20 h at 14°C. The molecular weight marker used was LW range PFG marker or λ DNA-HindIII digest, both from New England Biolabs. The probe used for Southern blot hybridization was a PCR product of 500 bp corresponding to neomycin or chloramphenicol resistance genes. Southern blots were performed with Hybond-N+ membranes as recommended by the manufacturer (GE Healthcare), and detection was done with the AlkPhos Direct Labeling kit (GE Healthcare).

## RESULTS

### Viral Replication and Recombination Proteins Are Responsible for the Generation of Plasmid Transducing Particles

To unravel the mechanisms that contribute to SPP1-mediated horizontal plasmid transfer we used *B. subtilis* BG214 strain, which is non-inducible for PBSX prophage and lacks prophage SPβ and the ICEBs1 integrative conjugative element. To analyze the role in antibiotic resistance transfer of SPP1 replication and recombination proteins, phages *sus34.1* and *sus36*, bearing mutations in genes 34.1 and 36 respectively, were constructed. SPP1 phage variants bearing mutations in the other genes were available in our phage collection (*sus19*, *sus53*, *sus109*, SPP1ΔA, tsB3). For comparison, a SPP1 *sus35* phage was also constructed, although a thermosensitive gene 35 mutant (the tsI20F phage) was available. The list of the bacteriophages used is shown in Table 2.

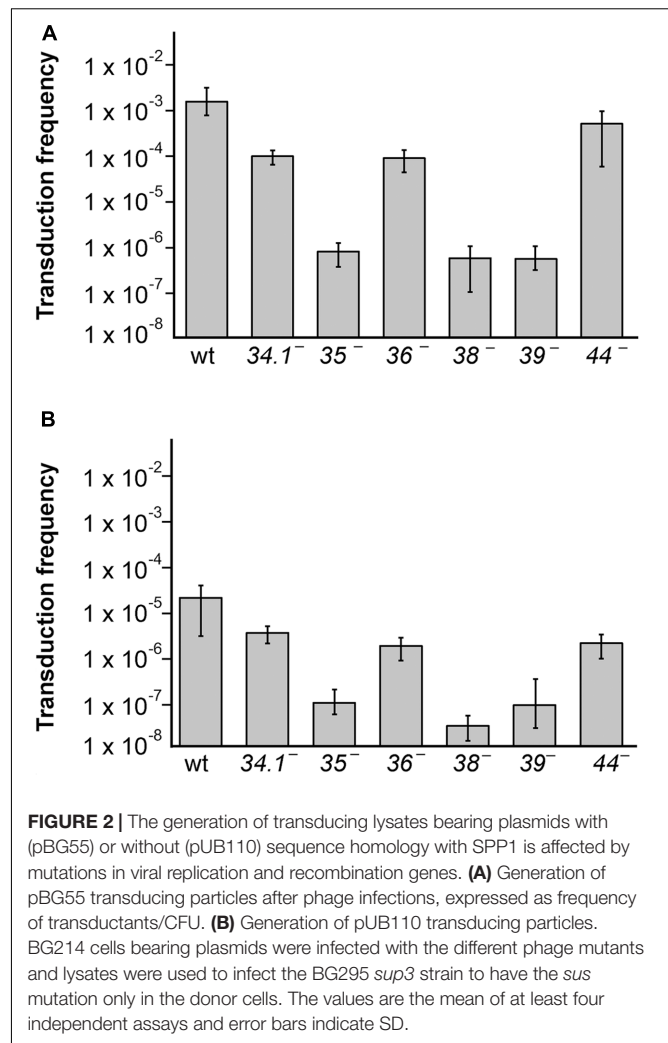
First we analyzed if G34.1P and G36P proteins, which were not yet studied *in vivo*, are essential for SPP1 replication (Figure 1). BG214 cells were grown until mid-exponential phase and then infected at MOI of 10 with the SPP1wt, SPP1ΔA, tsB3 (at restrictive temperature), or the different *sus* mutants (*sus34.1*, *sus35*, *sus36*, and *sus53*, a phage with a mutation in gene 39). After 90 min of infection, the phage lysates were collected and titrated. As previously observed, deletion of gene 44 reduced the



phage titer only 5-fold (Zecchi et al., 2012), whereas the mutation in gene 35, 38 or 39 completely abolished SPP1 amplification (Pedre et al., 1994; Ayora et al., 2002). The mutation in gene 36 reduced SPP1 titer only 6-fold, in agreement with the biochemical data showing that G36P can be replaced by the host SsbA during SPP1 DNA replication (Seco et al., 2013; Seco and Ayora, 2017). Deletion of the 34.1 gene reduced the phage titer 10-fold, and the size of the phage plaques was considerably smaller compared to the wt phage (Supplementary Figure S1). These results show that both, G36P, and G34.1P are not essential for phage amplification, although their defects reduce phage development.

To analyze if SPP1 replication and recombination proteins are involved in the generation of the transducing particle, the different *sus* mutant phages were used to infect BG214 cells bearing plasmid pBG55, a rolling circle replicating (RCR) plasmid with high-frequency of transduction (see Table 1 for more description). The lysates were collected after 90 min of infection, filtered and used to infect the BG295 *sup3* strain, to have the effect of phage *sus* mutation only in the donor and not in the recipient strain. The frequency of pBG55 transfer (Neomycin resistants [Nm<sup>R</sup>]/CFU) for the wt phage was similar to previously published results obtained using the BG214 strain, both as donor and as recipient (Deichelbohrer et al., 1985). These results show that the *sup3* genotype does not affect the transduction frequency. In parallel, infections with the thermosensitive phage mutants were performed at 50°C for 90 min. The lysates were then collected, filtered and used to infect BG214 cells at 30°C to have the effect of the thermosensitive mutation only in the donor, and not in the recipient strain. Mutations in genes 35, 38, or 39 blocked the transfer of the plasmid with homology (pBG55), with more than 1000-fold reduction in the transduction frequency (Figure 2A). A similar result was obtained with *sus109*, bearing a mutation in gene 40 (data not shown). Mutations in the exonuclease (G34.1P) or in the viral SSB (G36P) reduced the transduction frequency by ~12-fold, whereas the mutation in G44P only reduced it by ~4-fold.

To analyze if these proteins are also involved in the transfer of plasmids having no homology with the SPP1 phage, or

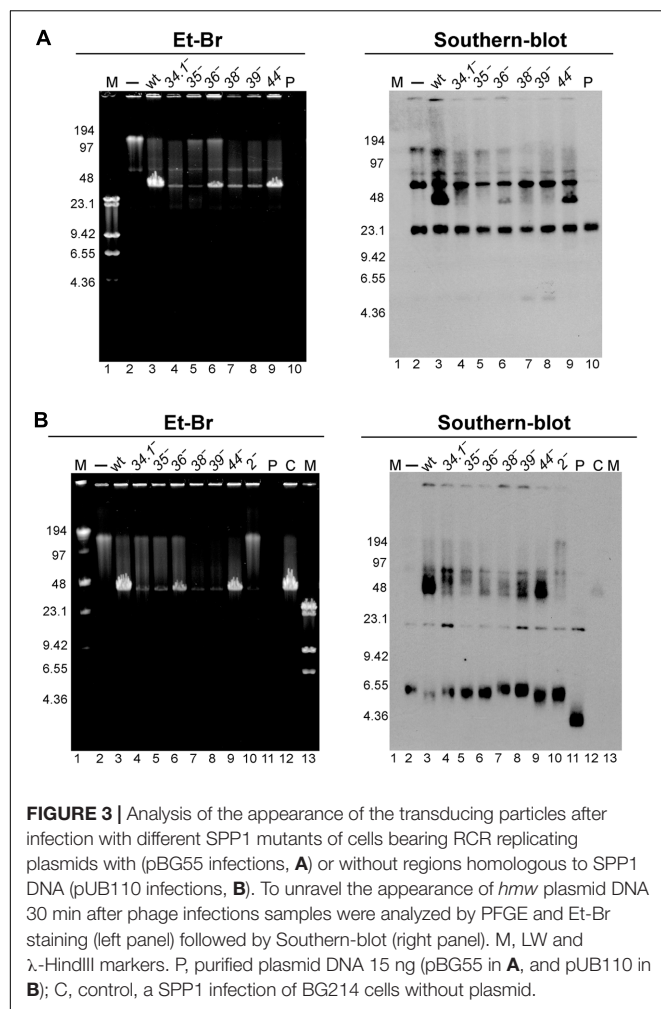


just very short homologous regions (sequences of 11–16 bp complementary to SPP1 DNA, see Supplementary Table S1) we performed transduction assays with the natural occurring pUB110 plasmid and the different phage mutants (Figure 2B). As already observed the transduction frequency of this plasmid was reduced by a factor of ~100-fold compared to the frequency of pBG55 transduction. The transduction frequencies were reduced in all of the SPP1 mutants, and similarly to the results obtained with the plasmid having homology, mutations in the recombinase or in replication proteins drastically reduced the phage-mediated transfer of pUB110, whereas mutations in the exonuclease, the SSB, or the HJ resolvase reduced the number of transductants/ml to a lesser extent.

## SPP1 Replication and Recombination Proteins Are Essential for the Generation of Plasmid Concatemeric DNA

Concatemeric plasmid DNA synthesis was observed with RCR plasmids after phage infection (Alonso et al., 1986; Bravo and Alonso, 1990). The results obtained in the previous section





**FIGURE 3 |** Analysis of the appearance of the transducing particles after infection with different SPP1 mutants of cells bearing RCR replicating plasmids with (pBG55 infections, **A**) or without regions homologous to SPP1 DNA (pUB110 infections, **B**). To unravel the appearance of *hmv* plasmid DNA 30 min after phage infections samples were analyzed by PFGE and Et-Br staining (left panel) followed by Southern-blot (right panel). M, LW and  $\lambda$ -HindIII markers. P, purified plasmid DNA 15 ng (pBG55 in **A**, and pUB110 in **B**); C, control, a SPP1 infection of BG214 cells without plasmid.

suggest that the essential viral recombination (G35P) and replication (G38P, G39P, and G40P) proteins could be responsible for the generation of this linear concatemeric plasmid DNA. To test this, we infected BG214 cells bearing pBG55 with the different phage mutants. After 30 min of infection, the infected cells were collected, total DNA was extracted, and separated by PFGE and Southern blotted to detect the production of concatemeric plasmid DNA forms. After infection with the wt phage the appearance of plasmid DNA that migrates with the bulk of SPP1 DNA (i.e., a multimeric plasmid DNA band of 44-kb) was observed (Figure 3A). In the absence of G35P, G38P or G39P, the production of this concatemeric band was not observed, consistent with the above result that mutations in these proteins block plasmid transduction. In agreement with its minor role in plasmid transfer, the 44-kb plasmid DNA band was observed after infection with phages bearing mutations in G34.1P, G36P, or in G44P. Moreover, the amount of 44-kb pBG55 DNA observed by PFGE and Southern blot correlated in these mutants with their transduction frequencies.

We also observed the appearance of a similar 44-kb plasmid band after infection with the wt SPP1 phage of cells bearing the natural pUB110 plasmid (Figure 3B). In concordance with

observations using the plasmid with extensive homology, the appearance of this 44-kb plasmid DNA band was clearly observed after infections with SPP1wt and SPP1 $\Delta$ A phages, which showed the highest transduction frequencies.

## Viral Replication and Recombination Proteins Are Not Involved in the Establishment of the Transduced Plasmid

The results presented above and in earlier reports (Deichelbohrer et al., 1985; Bravo and Alonso, 1990) indicate that a concatemeric  $\sim$ 44-kb plasmid DNA is encapsidated into the viral capsids. Once this concatemeric plasmid DNA (5.4 plasmid copies in the case of pBG55 plasmid) is injected into a recipient cell, it needs to circularize and monomerize to prepare the plasmid for correct replication and segregation cycles. The duplicated regions present in the concatemer could be used for monomerization, through a homologous recombination event, as it occurs during natural plasmid transformation (Kidane et al., 2009). In order to analyze if the viral replication and recombination machinery is involved in this monomerization and plasmid establishment process, we performed transduction assays with *sup3* as donor and wt as recipient cells (Table 3). It appeared that none of the viral proteins were required for the establishment of the transduced plasmid in the recipient cells.

## The Influence of Plasmid Copy Number and Replication Mode in Transduction

Plasmid-borne genes are transduced at much higher frequency than chromosomal-borne genes (Ferrari et al., 1978; Deichelbohrer et al., 1985), suggesting that copy number of plasmids could account for such differences. However, there is no tight correlation. As an example, it was published that the transduction frequency of plasmid pUB110, which has  $\sim$ 50 copies per cell (Viret and Alonso, 1988) is lower than that of pC1943 with  $\sim$ 15 copies per cell (Deichelbohrer et al., 1985). We confirmed these results (Table 4). This suggests that plasmid copy number is not the major determining factor, or not the only one. Other factors such as the presence of pseudo-*pac* sites, or of single-stranded (ssDNA) plasmid forms (recombinogenic particles, see below) could be the cause of this increased transduction frequency. Both plasmids, pUB110 and pC194, are RCR plasmids, but it was found that pC194 is more prone to formation of ssDNA than pUB110 (te Riele et al., 1986; Viret and Alonso, 1987).

To elucidate the influence of copy number, we compared the transduction efficiency of plasmid pUB110 ( $48 \pm 4$  copies/cell) and its derivative pUB110-cop1 ( $9 \pm 1$  copies/cell). pUB110-cop1 results from a single mutation in pUB110 plasmid, and therefore it has the same amount of ssDNA as the parental plasmid, but its copy number is reduced by 5-fold (Leonhardt, 1990). Both plasmids should have similar rates of circularization and establishment when they are injected into the recipient cell. As shown in Table 4, the transduction efficiency of pUB110-cop1 was proportionally reduced 4.6 times. In parallel we compared also the transduction frequencies of two other plasmids that

accumulate ssDNA, pC194 ( $15 \pm 2$  copies per cell, Alonso and Trautner, 1985) and pHP13 (a pTA1060 derivative,  $7 \pm 2$  copies per cell, Wang et al., 2004). Here also the transduction efficiency decreased by lowering the copy number of the plasmids. Nevertheless in all cases the transduction frequencies were higher for the plasmids accumulating ssDNA intermediates (Table 4).

Previous studies of the plasmid transduction by the SPP1 phage were done only with RCR plasmids. To determine the transduction frequency of theta replicating (TR) plasmids we used two such plasmids: pBT233 and pNDH33, which have a copy number similar to that of pHP13 plasmid (Table 1). Plasmid pBT233 is a pSM19035 derivative (erythromycin resistant), which has a copy number of  $\sim 8 \pm 2$ , and replicates unidirectionally by a DNA polymerase I (PolI)-dependent theta mechanism (Ceglowski et al., 1993a,b,c). Plasmid pNDH33 is a derivative of pBS72 (chloramphenicol resistant) with a copy number of  $\sim 6 \pm 1$  plasmids/cell (Nguyen et al., 2005; Phan et al., 2006). pNDH33 is thought to replicate by a DnaA-dependent and DNA PolI-independent theta type mechanism (Titok et al., 2003; Schumann, 2007). To compare TR and RCR plasmids, and to eliminate any resistance marker effects, the neomycin gene of the pUB110 was cloned into plasmid pBT233, to render plasmid pBT233N. The transduction frequency of the TR plasmid pBT233N was about 70-fold lower than that of pHP13. We measured also the transduction frequency of the second TR plasmid, pNDH33. This appeared to be also low, but only  $\sim 10$ -fold lower than that of pHP13 plasmid (Table 4). This higher transduction could be due to the occasional presence

in the pNDH33 plasmid of a *pseudo-pac* site or because of a 16 bp stretch of homology (Table 4 and Supplementary Table S1). When analyzing the fate of TR plasmids in infected cells, it was observed that, as with RCR plasmids, the infection with wt SPP1 phage produced the accumulation of a 44-kb plasmid DNA band, which was not observed after infection with a *sus35* mutant (Figure 4).

## The Presence of Homology to Phage Enhances the Transduction of TR Plasmids

When the phage packaging signal (*pac*) was integrated into the host chromosome, SPP1 mediated the transduction of chromosomal genes located close to the region of integration of the *pac* signal (Bravo et al., 1990). It was not tested if the presence of other SPP1 regions also increases the transduction frequencies of chromosomal DNA. To test this, we used the pBT233N derivative conferring  $\text{Nm}^R$ , which replicates via the theta-type mechanism as the chromosome. Different regions of SPP1 were cloned into pBT233N in order to evaluate whether the presence of *pac* sequence or the replication origin (*oriL*) results in higher transduction than simply homology to the phage (Table 5). Overall, the presence of a homologous region increased the transduction frequency of pBT233N plasmid by more than 1000-fold, and this increase was observed independently of the homologous region cloned (*pac*, *oriL*, or a 400 bp or 1000 bp region unrelated to replication and packaging processes).

**TABLE 3 |** Viral replication and recombination proteins are not involved in the establishment of transduced plasmids.

Donor strain <sup>a</sup> and plasmid	Recipient strain	Phage	Transduction Frequency <sup>b</sup>	SD <sup>c</sup>	TF <sup>M</sup> /TF <sup>wt</sup>
BG295 pBG55	BG214	SPP1 wt	$4.3 \times 10^{-3}$	$\pm 2.0 \times 10^{-3}$	1.00 <sup>d</sup>
BG295 pBG55	BG214	34.1 <sup>-</sup>	$3.4 \times 10^{-3}$	$\pm 1.5 \times 10^{-3}$	$7.9 \times 10^{-1}$
BG295 pBG55	BG214	35 <sup>-</sup>	$2.1 \times 10^{-3}$	$\pm 1.1 \times 10^{-3}$	$4.9 \times 10^{-1}$
BG295 pBG55	BG214	36 <sup>-</sup>	$3.4 \times 10^{-3}$	$\pm 1.8 \times 10^{-3}$	$7.9 \times 10^{-1}$
BG214 pBG55 <sup>e</sup>	BG214	38 <sup>-</sup>	$4.3 \times 10^{-3}$	$\pm 2.1 \times 10^{-3}$	$1.0 \times 10^0$
BG295 pBG55	BG214	39 <sup>-</sup>	$2.9 \times 10^{-3}$	$\pm 2.0 \times 10^{-3}$	$6.7 \times 10^{-1}$

<sup>a</sup>BG214 is the wild type strain and BG295 is the isogenic *sup3* strain. <sup>b</sup>The transduction frequency ( $\text{Neo}^R/\text{CFU}$ ) is the average of at least three independent experiments.

<sup>c</sup>SD: standard deviation. <sup>d</sup>The frequency of pBG55 plasmid transduction with the phage mutants (TF<sup>M</sup>) with respect to the wt phage (TF<sup>wt</sup>) is presented. <sup>e</sup>The 38<sup>-</sup> mutant is a thermosensitive phage (tsB3), therefore the infection was done at permissive temperature (30°C) and the transduction at non-permissive temperature (50°C) to have the mutation only in the recipient strain.

**TABLE 4 |** Transduction frequency of theta and rolling circle replicating plasmids without sequence homology with SPP1.

Plasmid	Ab <sup>R</sup> marker	Replication mechanism <sup>a</sup>	Copy number <sup>b</sup>	ssDNA production <sup>c</sup>	<i>pseudo-pac</i> site <sup>d</sup>	Transduction Frequency <sup>e</sup>	CI <sub>0.95</sub> <sup>f</sup>
pUB110	Nm	RCR	50	+	—	$2.1 \times 10^{-5}$	$\pm 1.3 \times 10^{-5}$
pUB110-cop1	Nm	RCR	15	+	—	$4.3 \times 10^{-6}$	$\pm 2.2 \times 10^{-6}$
pC194	Cm	RCR	15	+++	—	$5.2 \times 10^{-5}$	$\pm 4.7 \times 10^{-5}$
pHP13	Cm	RCR	5	+++	—	$5.9 \times 10^{-6}$	$\pm 2.9 \times 10^{-6}$
pBT233N	Nm	TR	8	—	—	$8.2 \times 10^{-8}$	$\pm 7.1 \times 10^{-8}$
pNDH33	Cm	TR	6	—	1	$6.4 \times 10^{-7}$	$\pm 3.2 \times 10^{-7}$

<sup>a</sup>RCR, rolling circle replication; TR, theta replication. <sup>b</sup>Plasmid copy numbers were reported in the literature and are presented here for comparison. <sup>c</sup>ssDNA production was reported in the literature and is presented here for comparison. <sup>d</sup>The *pac* motif (5'-CTATTGCGG↓C-3') is absent in all of the plasmids. Here the presence of a shorter motif that we call *pseudo-pac* site 5'-TTGCGG↓CW-3' is indicated. <sup>e</sup>The transduction frequency (transductants/CFU) is the mean of at least five independent experiments.

<sup>f</sup>CI, confidence interval.

Similarly, cloning into a RCR plasmid (pHP13) one of the phage origins of replication of SPP1 did not further increase the transduction frequency (Supplementary Table S2). Using of other TR-type replicon, pNDH33, provided similar results (Table 5). Furthermore, the accumulation of the 44-kb plasmid band was higher in the TR plasmids derivatives having homology with the phage (Figure 4, and data not shown).

## DISCUSSION

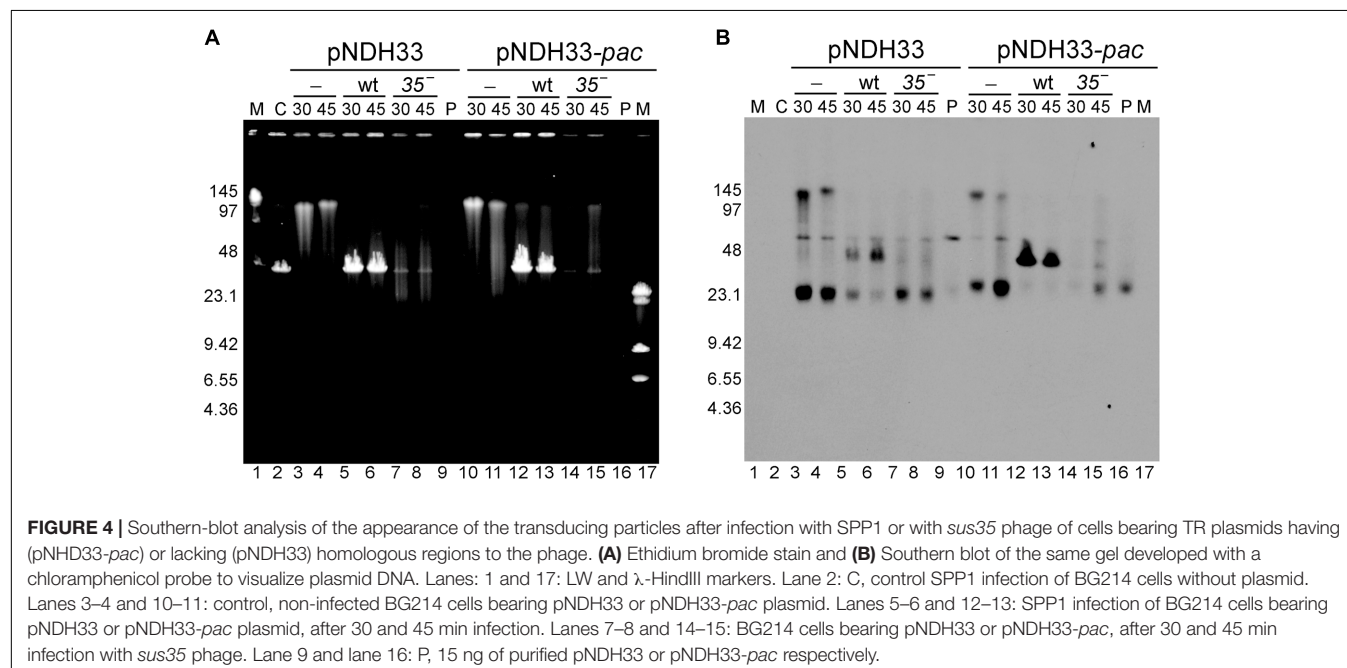
Until recently, it was thought that generalized transduction occurred at low frequency. However, recent single-cell analyses observed transduction rates close to 1% per plaque forming units when natural communities were used as recipients (Kenzaka et al., 2010). Therefore the study of the transduction mechanisms is essential to prevent this highly frequent horizontal gene transfer process, to avoid the spread of antibiotic resistance among bacteria. In this aspect, the SPP1 bacteriophage is a valuable model, because its replication, recombination,

and packaging machineries have been studied in deep for many years. Furthermore, it was recently reported that SPP1 can occasionally infect resistant cells when combined with sensitive cells, providing new routes for horizontal gene transfer (Tzipilevich et al., 2017). Previous biochemical studies assigned a role to SPP1 proteins G34.1P, G35P, G36P, G38P, G39P, G40P, and G44P in replication and recombination, but their contribution to generalized plasmid transduction remained unknown. Here we show that all SPP1 replication proteins contribute to horizontal plasmid transfer, although to a different extent. The origin binding protein (G38P), helicase loader (G39P), and helicase (G40P) are essential to produce concatemeric plasmid DNA, which is synthesized after phage infection. Infections with the *sus36* mutants show only a 10-fold reduction in the transduction frequency, probably due to potential complementation of the G36P function by cellular SsbA protein (Seco et al., 2013; Seco and Ayora, 2017). The SPP1 recombination proteins contribute to plasmid transfer to a different extent. The exonuclease G34.1P and the Holliday junction resolvase G44P only contribute partially to plasmid transduction, with a reduction of the

**TABLE 5 |** Transduction frequency of theta replicating plasmids bearing different SPP1 DNA regions.

Plasmid	Length of homologous region	Special features	Transduction Frequency <sup>a</sup>	CI <sub>0.95</sub> <sup>b</sup>
pBT233N	–	–	$8.2 \times 10^{-8}$	$\pm 7.1 \times 10^{-8}$
pBT233N-400	414 bp	homology	$3.1 \times 10^{-5}$	$\pm 1.9 \times 10^{-5}$
pBT233N-1300	1340 bp	homology	$4.5 \times 10^{-4}$	$\pm 3.0 \times 10^{-4}$
pBT233N- <i>pac</i>	412 bp	<i>pac</i>	$1.7 \times 10^{-4}$	$\pm 2.8 \times 10^{-4}$
pBT233N- <i>oriL</i>	360 bp	<i>oriL</i>	$1.5 \times 10^{-4}$	$\pm 1.6 \times 10^{-4}$
pNDH33	–	–	$5.6 \times 10^{-7}$	$\pm 2.9 \times 10^{-7}$
pNDH33-1300	1340 bp	homology	$2.0 \times 10^{-4}$	$\pm 8.5 \times 10^{-5}$
pNDH33- <i>pac</i>	412 bp	<i>pac</i>	$3.9 \times 10^{-4}$	$\pm 3.2 \times 10^{-4}$

<sup>a</sup>The transduction frequency (transductants/CFU) is the mean of at least five independent experiments. <sup>b</sup>CI, confidence interval.



transduction frequency of 12- and 5-fold in their mutants, respectively. The G35P recombinase is essential, with its inactivation leading to a >100-fold decrease.

Previous studies with SPP1 and RCR plasmids showed that: (i) the transduction of pUB110 and pC194 plasmids was enhanced 100- to 1000-fold when there was any homology between the plasmid and the SPP1 genome rather than with the specific *pac* signal; (ii) pUB110 and pC194 plasmid transduction was independent on RecA (Deichelbohrer et al., 1985), and (iii) linear plasmid concatemeric DNA (or high-molecular-weight [*hmw*] DNA) accumulated during phage infection, and in certain genetic backgrounds (Viret and Alonso, 1987; Viret et al., 1991). The synthesis of *hmw* DNA and its independence of the host-encoded recombinase (RecA) strongly suggests that the formation of transducing particles may rely on viral replication and/or recombination functions. In this work we show that the synthesis of this *hmw* DNA, and consequently transduction of RCR plasmids requires an active G35P protein. Biochemical analysis shows that G35P is an ATP-independent single-strand annealing enzyme, similar to the RecT enzyme encoded by the Rac prophage (Ayora et al., 2002). Both, G35P and RecT, belong to the Red $\beta$  family of viral single strand annealing proteins. To date, five different single strand annealing recombinase families have been identified in phages: Sak, Red $\beta$ , Erf, Sak4 and Gp2.5 (Lopes et al., 2010). These recombinases have gained increased attention in recent years because of their abundance in phage genomes (Lopes et al., 2010; Delattre et al., 2016), and also due to their wide use in recombineering systems (Datta et al., 2008; Sun et al., 2015). Many of these recombinases, including G35P, are essential for the phage life cycle (Zecchi et al., 2012; Neamah et al., 2017).

In this work we found that variations in copy-number affect the transduction frequency. Since the transduction is a stochastic process, it is expected that the more plasmid DNA in the cell the more generalized transducing phage particles should carry a plasmid copy and therefore the chances of transduction increase. The plasmids replicating in *B. subtilis* cells are either of the TR (circle-to-circle) type or RCR (sigma) type, and the products of both replication modes are usually covalently closed circular monomers (Khan, 2005). Comparing plasmids with similar copy number we observed that the frequency of transduction for RCR plasmids is ~60-fold higher than that for TR plasmids. This result suggests that the type of DNA replication also determines the transduction frequency. In the small RCR plasmids leading and lagging strand replication are uncoupled, and they contain two modules: the Rep protein with its cognate double-strand origin (DSO), and a single strand origin (SSO), which functions as the major initiation site for lagging-strand synthesis (Alonso et al., 1988; Espinosa et al., 1995; Khan, 2005). All RCR plasmids accumulate ssDNA although to a different extent: pUB110 accumulates traces and pC194 accumulates circular ssDNA (te Riele et al., 1986; Viret and Alonso, 1988). In contrast, the large low-copy-number TR plasmids, such as pBT233, which replicates via an unidirectional mechanism, do not accumulate circular ssDNA intermediates (Ceglowski et al., 1993b,c). We propose that the high transfer frequencies of some RCR plasmids may be correlated with the high accumulation of recombinogenic ssDNA

intermediates in these plasmids. Such ssDNA intermediates may constitute the substrates for formation of the transducing particles, through a recombination catalyzed by the G35P protein. This is in agreement with recent results observed with viral recombinases: when analyzing their recombineering activity *in vivo*, it was found that they catalyze single-strand annealing preferentially on the lagging strand (van Kessel and Hatfull, 2008; Mosberg et al., 2010; Lajoie et al., 2012; Fricker and Peters, 2014; Ander et al., 2015). We propose that all the phages encoding recombinases will transduce RCR plasmids with high efficiency by the mechanism of viral recombinase-mediated generalized transduction. Furthermore, we also observed that the transduction of the pUB110 and pNDH33 plasmids, which do not have an extensive region of homology, was strongly reduced in infections with the *sus35* mutant (Figures 2, 4). All phage recombinases studied so far are single-strand annealing proteins that promote genetic recombination under more permissive conditions than RecA (Scaltriti et al., 2011; De Paepe et al., 2014; Menouni et al., 2015). Our results suggest that G35P contributes to the transfer of natural plasmids by catalyzing a recombination reaction using small stretches of homology found in many plasmids (Supplementary Table S1).

The different contributions of the SPP1 recombination proteins to plasmid transduction, together with the high recombinogenic nature of the RCR plasmids, suggest that the initial DNA substrate, used for the production of transducing particles by recombination, is indeed ssDNA. This is consistent with the result that the G34.IP exonuclease, which resects the dsDNA ends to generate the appropriate substrate for the recombinase (Martinez-Jimenez et al., 2005), has a minor role in plasmid transfer. Similarly, we found that the SPP1 SSB protein, G36P, only slightly contributes to the mechanisms of plasmid transduction. However, in some phages the recombinases require the activity of their cognate SSB proteins to perform their function (Neamah et al., 2017).

It was previously observed with RCR plasmids that any SPP1 DNA segment larger than 50 bp, cloned into such plasmids, greatly increased the transduction frequency (Deichelbohrer et al., 1985; Alonso et al., 1986). We extend this observation to TR plasmids, where the transduction frequency was highly increased, independently of what is the region of homology cloned, whether it was the packaging sequence, a phage origin of replication, or any other region of homology. Similarly, the cloning of the origin of replication of SPP1 (*oriL*) into a RCR-type plasmid did not further increase its transduction frequency (Supplementary Table S2). We conclude that any DNA region homologous to the phage genome increases the frequency of horizontal transfer of plasmids, independently of their replication mechanism. Enhanced transduction of plasmids bearing homology with phage DNA has been also observed with phage T4, which codes for a different recombinase, the UvsX protein (Kreuzer et al., 1988), and with *Salmonella typhimurium* phage P22, which codes for the Erf recombinase (Orbach and Jackson, 1982).

How is the plasmid substrate for generalized transduction generated? Three different mechanisms could account for the generation of a concatemeric plasmid DNA with high frequency



of transduction. In the first model, the multiple tandem repeats of plasmid DNA might be produced by intermolecular recombination, as proposed for P22 plasmid transduction (Mann and Slauch, 1997). This mechanism resembles phage T4 generation of concatemeric DNA during its replication (Kreuzer, 2000; Mosig et al., 2001). Here, multiple strand invasions catalyzed by the ATP-dependent RecA-like recombinase encoded by this phage, UvsX, and the resolution of the Holliday junction intermediates by its Holliday junction resolvase Gp49 (also called EndoVII), produce the concatemeric DNA, as well as the transducing particle (Kreuzer et al., 1988; Kreuzer, 2000; Mosig et al., 2001). We do not favor this hypothesis in the SPP1 system, because we found that the Holliday junction resolvase G44P has only a minor role in plasmid pBG55 and pUB110 transduction. In the second model, plasmid over-replication leads to the accumulation of linear concatemeric *hmw* DNA (Cohen and Clark, 1986; Viret and Alonso, 1987; Viret et al., 1991). The accumulation of linear head-to-tail multigenome-length plasmid DNA (*hmw* DNA) in the absence of RecBCD/AddAB was documented in both *Escherichia coli* and *B. subtilis* cells (Silberstein and Cohen, 1987; Viret and Alonso, 1987). Indeed, upon infection, many bacteriophages directly or indirectly inactivate end-resection catalyzed by this host encoded multi-subunit helicase-nuclease enzyme (Szczepanska, 2009). It was observed that the synthesis of pC194 or pUB110 *hmw* plasmid DNA occurred in the absence of plasmid-encoded Rep protein, and required DNA PolI, RecA and pre-primosomal proteins (e.g., DnaB) (Viret and Alonso, 1987; Leonhardt et al., 1991; Viret et al., 1991). Analysis of this *hmw* plasmid DNA by electron microscopy displayed linear DNA molecules up to 100 kb in size, which were either single-stranded, double-stranded or duplex DNA with single-stranded tailed ends (Leonhardt et al., 1991). This *hmw* DNA can be encapsidated into a viral prohead by a headful packaging mechanism (Schmidt and Schmieger, 1984; Schmieger, 1984). If this model is correct, the presence of a *pac* signal will significantly increase the encapsidation of the plasmid *hmw* DNA, and we found that there was not an increase in the transduction frequency when plasmids contained the *pac* signal. In the third model, phage infection arrests host and plasmid replication. Then SPP1-dependent replication restarts, and the linear plasmid concatemer is synthesized. This is consistent with the result that the phage G38P protein may act as a PriA-like enzyme, restarting DNA replication outside form a replication origin (Seco et al., 2013; Seco and Ayora, 2017). In this *de novo* synthesis of plasmid DNA, a viral *pac* site might be gained by recombination and recognized by the viral packaging machinery (Alonso et al., 1986; Bravo et al., 1990; Viret et al., 1991). In this model, the phage might form a phage-plasmid chimera and the plasmid hijacks the viral replication machinery to promote *de novo* synthesis of linear plasmid concatemeric DNA. The concatemeric plasmid DNA is then packaged into an empty prohead by the headful mechanism, indistinguishable of viral DNA, provided that the packaged substrate is larger than mature phage DNA. Our data support the third model, because we found that in infections with a

phage bearing a mutation in the terminase (*sus19* infections), plasmid concatemers up to 200-kb long are produced after phage infection (Figure 3B and Supplementary Figure S2). This model explains also the requirement of viral replication proteins for the formation of the transducing particles. However, we were unable to detect the phage-plasmid chimeras, which might be rapidly processed to produce the plasmid head-to-tail concatemers.

Our results show that the establishment of the transduced concatemeric plasmid in the host is independent of phage encoded recombination functions, which only participate in the generation of the transducing particle. We propose that the injected linear concatemer can be converted into a circular form by the homologous recombination machinery of the recipient cells. In this respect, transduction of plasmids might have similar host requirements as the resolution of phage-plasmid chimeras analyzed in the P22 and SPP1 systems (Orbach and Jackson, 1982; Alonso et al., 1992). In the former case, the plasmid integrated into the phage genome has to be excised from the genome of the defective phage prior to establishment, whereas in the latter case the head-to-tail plasmid concatemer has to recombine intramolecularly to facilitate plasmid establishment. This process was found to be RecA-independent but dependent on host RecO and RecR functions that also catalyze single-strand annealing (Alonso et al., 1992; Manfredi et al., 2008).

## AUTHOR CONTRIBUTIONS

AV-R, ML-S, AQ-O, and SA: performed the experiments; AV-R, AS, and SA: analyzed data; SA: conceived the project, integrated the results and wrote the paper.

## FUNDING

This work was partially supported by Spanish grants BFU2012-39879-C02-02 and BFU2015-67065-P from MINECO to SA, and PathoBactEvol (ANR-12-ADAP-0018) from ANR to AS.

## ACKNOWLEDGMENTS

We thank J. C. Alonso (CNB-CSIC, Spain) for providing us with pUB110 and pHP13 plasmid derivatives, and for critically reading this manuscript. Plasmid pNDH33 was kindly provided by Wolfgang Schumann (University of Bayreuth, Germany).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01816/full#supplementary-material>

## REFERENCES

- Alonso, J. C., Luder, G., and Trautner, T. A. (1986). Requirements for the formation of plasmid-transducing particles of *Bacillus subtilis* bacteriophage SPP1. *EMBO J.* 5, 3723–3728.
- Alonso, J. C., Luder, G., and Trautner, T. A. (1992). Intramolecular homologous recombination in *Bacillus subtilis* 168. *Mol. Gen. Genet.* 236, 60–64.
- Alonso, J. C., Stiege, C. A., Taylor, R. H., and Viret, J. F. (1988). Functional analysis of the dna (Ts) mutants of *Bacillus subtilis*: plasmid pUB110 replication as a model system. *Mol. Gen. Genet.* 214, 482–489. doi: 10.1007/BF00330484
- Alonso, J. C., Tavares, P., Lurz, R., and Trautner, T. A. (2006). “Bacteriophage SPP1,” in *The Bacteriophages*, ed. R. Calendar (New York, NY: Oxford University Press), 18.
- Alonso, J. C., and Trautner, T. A. (1985). A gene controlling segregation of the *Bacillus subtilis* plasmid pC194. *Mol. Gen. Genet.* 198, 427–431. doi: 10.1007/BF00332934
- Ander, M., Subramaniam, S., Fahmy, K., Stewart, A. F., and Schaffer, E. (2015). A single-strand annealing protein clamps DNA to detect and secure homology. *PLoS Biol.* 13:e1002213. doi: 10.1371/journal.pbio.1002213
- Ayora, S., Missich, R., Mesa, P., Lurz, R., Yang, S., Egelman, E. H., et al. (2002). Homologous-pairing activity of the *Bacillus subtilis* bacteriophage SPP1 replication protein G35P. *J. Biol. Chem.* 277, 35969–35979. doi: 10.1074/jbc.M20446720
- Ayora, S., Stasiak, A., and Alonso, J. C. (1999). The *Bacillus subtilis* bacteriophage SPP1 G35P delivers and activates the G40P DNA helicase upon interacting with the G38P-bound replication origin. *J. Mol. Biol.* 288, 71–85. doi: 10.1006/jmbi.1999.2662
- Bravo, A., and Alonso, J. C. (1990). The generation of concatemeric plasmid DNA in *Bacillus subtilis* as a consequence of bacteriophage SPP1 infection. *Nucleic Acids Res.* 18, 4651–4657. doi: 10.1093/nar/18.16.4651
- Bravo, A., Alonso, J. C., and Trautner, T. A. (1990). Functional analysis of the *Bacillus subtilis* bacteriophage SPP1 pac site. *Nucleic Acids Res.* 18, 2881–2886. doi: 10.1093/nar/18.10.2881
- Brown-Jaque, M., Calero-Caceres, W., and Muniesa, M. (2015). Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid* 79, 1–7. doi: 10.1016/j.plasmid.2015.01.001
- Brussow, H., Canchaya, C., and Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi: 10.1128/MMBR.68.3.560-602.2004
- Burger, K. J., and Trautner, T. A. (1978). Specific labelling of replicating SPP1 DNA: analysis of viral DNA synthesis and identification of phage DNA-genes. *Mol. Gen. Genet.* 166, 277–285.
- Camacho, A. G., Gual, A., Lurz, R., Tavares, P., and Alonso, J. C. (2003). *Bacillus subtilis* bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. *J. Biol. Chem.* 278, 23251–23259. doi: 10.1074/jbc.M301805200
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L., and Brussow, H. (2003). Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* 6, 417–424. doi: 10.1016/S1369-5274(03)00086-9
- Canosi, U., Luder, G., and Trautner, T. A. (1982). SPP1-mediated plasmid transduction. *J. Virol.* 44, 431–436.
- Ceglowski, P., Boitsov, A., Chai, S., and Alonso, J. C. (1993a). Analysis of the stabilization system of pSM19035-derived plasmid pBT233 in *Bacillus subtilis*. *Gene* 136, 1–12.
- Ceglowski, P., Boitsov, A., Karamyan, N., Chai, S., and Alonso, J. C. (1993b). Characterization of the effectors required for stable inheritance of *Streptococcus pyogenes* pSM19035-derived plasmids in *Bacillus subtilis*. *Mol. Gen. Genet.* 241, 579–585.
- Ceglowski, P., Lurz, R., and Alonso, J. C. (1993c). Functional analysis of pSM19035-derived replicons in *Bacillus subtilis*. *FEMS Microbiol. Lett.* 109, 145–150.
- Chai, S., Bravo, A., Luder, G., Nedlin, A., Trautner, T. A., and Alonso, J. C. (1992). Molecular analysis of the *Bacillus subtilis* bacteriophage SPP1 region encompassing genes 1 to 6. The products of gene 1 and gene 2 are required for pac cleavage. *J. Mol. Biol.* 224, 87–102. doi: 10.1016/0022-2836(92)90578-8
- Chai, S., Lurz, R., and Alonso, J. C. (1995). The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J. Mol. Biol.* 252, 386–398. doi: 10.1006/jmbi.1995.0505
- Chai, S., Szepan, U., and Alonso, J. C. (1997). *Bacillus subtilis* bacteriophage SPP1 terminase has a dual activity: it is required for the packaging initiation and represses its own synthesis. *Gene* 184, 251–256. doi: 10.1016/S0378-1119(96)00609-9
- Chai, S., Szepan, U., Luder, G., Trautner, T. A., and Alonso, J. C. (1993). Sequence analysis of the left end of the *Bacillus subtilis* bacteriophage SPP1 genome. *Gene* 129, 41–49. doi: 10.1016/0378-1119(93)90694-X
- Cohen, A., and Clark, A. J. (1986). Synthesis of linear plasmid multimers in *Escherichia coli* K-12. *J. Bacteriol.* 167, 327–335. doi: 10.1128/jb.167.1.327-335.1986
- Datta, S., Costantino, N., Zhou, X., and Court, D. L. (2008). Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1626–1631. doi: 10.1073/pnas.0709089105
- De Paepe, M., Hutinet, G., Son, O., Amarir-Bouhram, J., Schbath, S., and Petit, M. A. (2014). Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet.* 10:e1004181. doi: 10.1371/journal.pgen.1004181
- Deichelbohrer, I., Alonso, J. C., Luder, G., and Trautner, T. A. (1985). Plasmid transduction by *Bacillus subtilis* bacteriophage SPP1: effects of DNA homology between plasmid and bacteriophage. *J. Bacteriol.* 162, 1238–1243.
- Delattre, H., Souiai, O., Fagoonee, K., Guerois, R., and Petit, M. A. (2016). Phagonaut: a web-based interface for phage synteny browsing and protein function prediction. *Virology* 496, 42–50. doi: 10.1016/j.virol.2016.05.007
- Espinosa, M., del Solar, G., Rojo, F., and Alonso, J. C. (1995). Plasmid rolling circle replication and its control. *FEMS Microbiol. Lett.* 130, 111–120. doi: 10.1111/j.1574-6968.1995.tb07707.x
- Ferrari, E., Canosi, U., Galizzi, A., and Mazza, G. (1978). Studies on transduction process by SPP1 phage. *J. Gen. Virol.* 41, 563–572. doi: 10.1099/0022-1317-41-3-563
- Fricker, A. D., and Peters, J. E. (2014). Vulnerabilities on the lagging-strand template: opportunities for mobile elements. *Annu. Rev. Genet.* 48, 167–186. doi: 10.1146/annurev-genet-120213-092046
- Giovanetti, E., Brenciani, A., Morroni, G., Tiberi, E., Pasquaroli, S., Mingoia, M., et al. (2014). Transduction of the *Streptococcus pyogenes* bacteriophage Phm46.1, carrying resistance genes *mef(A)* and *tet(O)*, to other *Streptococcus* species. *Front. Microbiol.* 5:746. doi: 10.3389/fmicb.2014.00746
- Gual, A., Camacho, A. G., and Alonso, J. C. (2000). Functional analysis of the terminase large subunit, G2P, of *Bacillus subtilis* bacteriophage SPP1. *J. Biol. Chem.* 275, 35311–35319. doi: 10.1074/jbc.M004309200
- Haima, P., Bron, S., and Venema, G. (1987). The effect of restriction on shotgun cloning and plasmid stability in *Bacillus subtilis* Marburg. *Mol. Gen. Genet.* 209, 335–342. doi: 10.1007/BF00329663
- Horinouchi, S., and Weisblum, B. (1982). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* 150, 815–825.
- Iked, H., and Tomizawa, J. I. (1965). Transducing fragments in generalized transduction by phage P1. I. Molecular origin of the fragments. *J. Mol. Biol.* 14, 85–109. doi: 10.1016/S0022-2836(65)80232-7
- Kenzaka, T., Tani, K., and Nasu, M. (2010). High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME J.* 4, 648–659. doi: 10.1038/ismej.2009.145
- Khan, S. A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* 53, 126–136. doi: 10.1016/j.plasmid.2004.12.008
- Kidane, D., Carrasco, B., Manfredi, C., Rothmaier, K., Ayora, S., Tadesse, S., et al. (2009). Evidence for different pathways during horizontal gene transfer in competent *Bacillus subtilis* cells. *PLoS Genet.* 5:e1000630. doi: 10.1371/journal.pgen.1000630
- Kreuzer, K. N. (2000). Recombination-dependent DNA replication in phage T4. *Trends Biochem. Sci.* 25, 165–173. doi: 10.1016/S0968-0004(00)01559-0
- Kreuzer, K. N., Yap, W. Y., Menkens, A. E., and Engman, H. W. (1988). Recombination-dependent replication of plasmids during bacteriophage T4 infection. *J. Biol. Chem.* 263, 11366–11373.
- Lajoie, M. J., Gregg, C. J., Mosberg, J. A., Washington, G. C., and Church, G. M. (2012). Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res.* 40:e170. doi: 10.1093/nar/gks751

- Lennox, E. S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* 1, 190–206. doi: 10.1016/0042-6822(55)90016-7
- Leonhardt, H. (1990). Identification of a low-copy-number mutation within the pUB110 replicon and its effect on plasmid stability in *Bacillus subtilis*. *Gene* 94, 121–124. doi: 10.1016/0378-1119(90)90477-9
- Leonhardt, H., Lurz, R., and Alonso, J. C. (1991). Physical and biochemical characterization of recombination-dependent synthesis of linear plasmid multimers in *Bacillus subtilis*. *Nucleic Acids Res.* 19, 497–503. doi: 10.1093/nar/19.3.497
- Lo, Piano, A., Martinez-Jimenez, M. I., Zecchi, L., and Ayora, S. (2011). Recombination-dependent concatemeric viral DNA replication. *Virus Res.* 160, 1–14. doi: 10.1016/j.virusres.2011.06.009
- Lopes, A., Amarir-Bouhram, J., Faure, G., Petit, M. A., and Guerois, R. (2010). Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res.* 38, 3952–3962. doi: 10.1093/nar/gkq096
- Manfredi, C., Carrasco, B., Ayora, S., and Alonso, J. C. (2008). *Bacillus subtilis* RecO nucleates RecA onto SsbA-coated single-stranded DNA. *J. Biol. Chem.* 283, 24837–24847. doi: 10.1074/jbc.M802002200
- Mann, B. A., and Schlauch, J. M. (1997). Transduction of low-copy number plasmids by bacteriophage P22. *Genetics* 146, 447–456.
- Martinez-Jimenez, M. I., Alonso, J. C., and Ayora, S. (2005). *Bacillus subtilis* bacteriophage SPP1-encoded gene 34.1 product is a recombination-dependent DNA replication protein. *J. Mol. Biol.* 351, 1007–1019. doi: 10.1016/j.jmb.2005.06.064
- Martinez-Jimenez, M. I., Mesa, P., and Alonso, J. C. (2002). *Bacillus subtilis* tau subunit of DNA polymerase III interacts with bacteriophage SPP1 replicative DNA helicase G40P. *Nucleic Acids Res.* 30, 5056–5064. doi: 10.1093/nar/gkf650
- Maslanova, I., Doskar, J., Varga, M., Kuntova, L., Muzik, J., Maluskova, D., et al. (2013). Bacteriophages of *Staphylococcus aureus* efficiently package various bacterial genes and mobile genetic elements including SCCmec with different frequencies. *Environ. Microbiol. Rep.* 5, 66–73. doi: 10.1111/j.1758-2229.2012.00378.x
- Menouni, R., Hutinet, G., Petit, M. A., and Ansaldi, M. (2015). Bacterial genome remodeling through bacteriophage recombination. *FEMS Microbiol. Lett.* 362, 1–10. doi: 10.1093/femsle/fnu022
- Missich, R., Weise, F., Chai, S., Lurz, R., Pedre, X., and Alonso, J. C. (1997). The replisome organizer (G38P) of *Bacillus subtilis* bacteriophage SPP1 forms specialized nucleoprotein complexes with two discrete distant regions of the SPP1 genome. *J. Mol. Biol.* 270, 50–64. doi: 10.1006/jmbi.1997.1060
- Mosberg, J. A., Lajoie, M. J., and Church, G. M. (2010). Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* 186, 791–799. doi: 10.1534/genetics.110.120782
- Mosig, G., Gewin, J., Luder, A., Colowick, N., and Vo, D. (2001). Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8306–8311. doi: 10.1073/pnas.131007398
- Neamah, M. M., Mir-Sanchis, I., Lopez-Sanz, M., Acosta, S., Baquedano, I., Haag, A. F., et al. (2017). Sak and Sak4 recombinases are required for bacteriophage replication in *Staphylococcus aureus*. *Nucleic Acids Res.* 45, 6507–6519. doi: 10.1093/nar/gkx308
- Nguyen, H. D., Nguyen, Q. A., Ferreira, R. C., Ferreira, L. C., Tran, L. T., and Schumann, W. (2005). Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. *Plasmid* 54, 241–248. doi: 10.1016/j.plasmid.2005.05.001
- Oliveira, L., Alonso, J. C., and Tavares, P. (2005). A defined in vitro system for DNA packaging by the bacteriophage SPP1: insights into the headful packaging mechanism. *J. Mol. Biol.* 353, 529–539. doi: 10.1016/j.jmb.2005.08.063
- Oliveira, L., Tavares, P., and Alonso, J. C. (2013). Headful DNA packaging: bacteriophage SPP1 as a model system. *Virus Res.* 173, 247–259. doi: 10.1016/j.virusres.2013.01.021
- Orbach, M. J., and Jackson, E. N. (1982). Transfer of chimeric plasmids among *Salmonella typhimurium* strains by P22 transduction. *J. Bacteriol.* 149, 985–994.
- Pedre, X., Weise, F., Chai, S., Luder, G., and Alonso, J. C. (1994). Analysis of cis and trans acting elements required for the initiation of DNA replication in the *Bacillus subtilis* bacteriophage SPP1. *J. Mol. Biol.* 236, 1324–1340. doi: 10.1016/0022-2836(94)90061-2
- Penades, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R. P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* 23, 171–178. doi: 10.1016/j.mib.2014.11.019
- Phan, T. T., Nguyen, H. D., and Schumann, W. (2006). Novel plasmid-based expression vectors for intra- and extracellular production of recombinant proteins in *Bacillus subtilis*. *Protein Expr. Purif.* 46, 189–195. doi: 10.1016/j.pep.2005.07.005
- Ruzin, A., Lindsay, J., and Novick, R. P. (2001). Molecular genetics of SaPI1—a mobile pathogenicity island in *Staphylococcus aureus*. *Mol. Microbiol.* 41, 365–377. doi: 10.1046/j.1365-2958.2001.02488.x
- Scaltriti, E., Launay, H., Genois, M. M., Bron, P., Rivetti, C., Grolli, S., et al. (2011). Lactococcal phage p2 ORF35-Sak3 is an ATPase involved in DNA recombination and AbiK mechanism. *Mol. Microbiol.* 80, 102–116. doi: 10.1111/j.1365-2958.2011.07561.x
- Schmidt, C., and Schmieger, H. (1984). Selective transduction of recombinant plasmids with cloned pac sites by *Salmonella* phage P22. *Mol. Gen. Genet.* 196, 123–128. doi: 10.1007/BF00334103
- Schmieger, H. (1984). pac sites are indispensable for in vivo packaging of DNA by phage P22. *Mol. Gen. Genet.* 195, 252–255. doi: 10.1007/BF00332755
- Schumann, W. (2007). Production of recombinant proteins in *Bacillus subtilis*. *Adv. Appl. Microbiol.* 62, 137–189. doi: 10.1016/S0065-2164(07)62006-1
- Seco, E. M., and Ayora, S. (2017). *Bacillus subtilis* DNA polymerases, PolC and DnaE, are required for both leading and lagging strand synthesis in SPP1 origin-dependent DNA replication. *Nucleic Acids Res.* 45, 8302–8313. doi: 10.1093/nar/gkx493
- Seco, E. M., Zinder, J. C., Manhart, C. M., Lo Piano, A., McHenry, C. S., and Ayora, S. (2013). Bacteriophage SPP1 DNA replication strategies promote viral and disable host replication in vitro. *Nucleic Acids Res.* 41, 1711–1721. doi: 10.1093/nar/gks1290
- Silberstein, Z., and Cohen, A. (1987). Synthesis of linear multimers of OriC and pBR322 derivatives in *Escherichia coli* K-12: role of recombination and replication functions. *J. Bacteriol.* 169, 3131–3137. doi: 10.1128/jb.169.7.3131-3137.1987
- Sun, Z., Deng, A., Hu, T., Wu, J., Sun, Q., Bai, H., et al. (2015). A high-efficiency recombineering system with PCR-based ssDNA in *Bacillus subtilis* mediated by the native phage recombinase GP35. *Appl. Microbiol. Biotechnol.* 99, 5151–5162. doi: 10.1007/s00253-015-6485-5
- Szczepanska, A. K. (2009). Bacteriophage-encoded functions engaged in initiation of homologous recombination events. *Crit. Rev. Microbiol.* 35, 197–220. doi: 10.1080/10408410902983129
- te Riele, H., Michel, B., and Ehrlich, S. D. (1986). Single-stranded plasmid DNA in *Bacillus subtilis* and *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* 83, 2541–2545. doi: 10.1073/pnas.83.8.2541
- Titok, M. A., Chapuis, J., Selezneva, Y. V., Lagodich, A. V., Prokulevich, V. A., Ehrlich, S. D., et al. (2003). *Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. *Plasmid* 49, 53–62. doi: 10.1016/S0147-619X(02)00109-9
- Touchon, M., Moura de Sousa, J. A., and Rocha, E. P. (2017). Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr. Opin. Microbiol.* 38, 66–73. doi: 10.1016/j.mib.2017.04.010
- Tzipilevich, E., Habusha, M., and Ben-Yehuda, S. (2017). Acquisition of phage sensitivity by bacteria through exchange of phage receptors. *Cell* 168, 186e–199e. doi: 10.1016/j.cell.2016.12.003
- van Kessel, J. C., and Hatfull, G. F. (2008). Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol. Microbiol.* 67, 1094–1107. doi: 10.1111/j.1365-2958.2008.06109.x
- Viret, J. F., and Alonso, J. C. (1987). Generation of linear multigenome-length plasmid molecules in *Bacillus subtilis*. *Nucleic Acids Res.* 15, 6349–6367. doi: 10.1093/nar/15.16.6349
- Viret, J. F., and Alonso, J. C. (1988). A DNA sequence outside the pUB110 minimal replicon is required for normal replication in *Bacillus subtilis*. *Nucleic Acids Res.* 16, 4389–4406. doi: 10.1093/nar/16.10.4389
- Viret, J. F., Bravo, A., and Alonso, J. C. (1991). Recombination-dependent concatemeric plasmid replication. *Microbiol. Rev.* 55, 675–683.

- Wang, J. D., Rokop, M. E., Barker, M. M., Hanson, N. R., and Grossman, A. D. (2004). Multicopy plasmids affect replisome positioning in *Bacillus subtilis*. *J. Bacteriol.* 186, 7084–7090. doi: 10.1128/JB.186.21.7084-7090.2004
- Winstel, V., Kuhner, P., Krismer, B., Peschel, A., and Rohde, H. (2015). Transfer of plasmid DNA to clinical coagulase-negative staphylococcal pathogens by using a unique bacteriophage. *Appl. Environ. Microbiol.* 81, 2481–2488. doi: 10.1128/AEM.04190-14
- Yasbin, R. E., and Young, F. E. (1974). Transduction in *Bacillus subtilis* by bacteriophage SPP1. *J. Virol.* 14, 1343–1348.
- Zecchi, L., Lo Piano, A., Suzuki, Y., Canas, C., Takeyasu, K., and Ayora, S. (2012). Characterization of the Holliday junction resolving enzyme encoded by the *Bacillus subtilis* bacteriophage SPP1. *PLOS ONE* 7:e48440. doi: 10.1371/journal.pone.0048440
- Zinder, N. D., and Lederberg, J. (1952). Genetic exchange in *Salmonella*. *J. Bacteriol.* 64, 679–699.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2017 Valero-Rello, López-Sanz, Quevedo-Olmos, Sorokin and Ayora. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Phages in the Human Body

Ferran Navarro<sup>1</sup> and Maite Muniesa<sup>2\*</sup>

<sup>1</sup> Servei de Microbiologia, Hospital de la Santa Creu i Sant Pau, Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain, <sup>2</sup> Department of Microbiology, University of Barcelona, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### Reviewed by:

Guillem Prats,  
Autonomous University of Barcelona,  
Spain

Steven P. T. Hooton,  
University of Nottingham, UK

Radosław Pluta,  
International Institute of Molecular  
and Cell Biology in Warsaw, Poland

### \*Correspondence:

Maite Muniesa  
mmuniesa@ub.edu

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 13 February 2017

**Accepted:** 20 March 2017

**Published:** 04 April 2017

### Citation:

Navarro F and Muniesa M (2017)  
Phages in the Human Body.  
Front. Microbiol. 8:566.  
doi: 10.3389/fmicb.2017.00566

Bacteriophages, viruses that infect bacteria, have re-emerged as powerful regulators of bacterial populations in natural ecosystems. Phages invade the human body, just as they do other natural environments, to such an extent that they are the most numerous group in the human virome. This was only revealed in recent metagenomic studies, despite the fact that the presence of phages in the human body was reported decades ago. The influence of the presence of phages in humans has yet to be evaluated; but as in marine environments, a clear role in the regulation of bacterial populations could be envisaged, that might have an impact on human health. Moreover, phages are excellent vehicles of genetic transfer, and they contribute to the evolution of bacterial cells in the human body by spreading and acquiring DNA horizontally. The abundance of phages in the human body does not pass unnoticed and the immune system reacts to them, although it is not clear to what extent. Finally, the presence of phages in human samples, which most of the time is not considered, can influence and bias microbiological and molecular results; and, in view of the evidences, some studies suggest that more attention needs to be paid to their interference.

**Keywords:** bacteriophages, human biomes, homeostasis, metagenomics, diagnosis, virome

## INTRODUCTION

Bacteriophages were discovered in the second decade of the 20th century (Twort, 1915; D'Herelle, 1917). It was initially suggested the idea they could be used to lyse pathogenic bacteria as a treatment of infectious diseases. However, the idea was rapidly abandoned in western countries due to the introduction of antibiotics. For decades, phages have been the most common model entities for the study of viruses and their replication cycles. Studies of certain model phages have contributed significantly to the advancement of molecular biology, for example in identifying the basis of genetic material, as the code of nucleotide triplets of individual amino acids (Crick et al., 1961), and the restriction enzymes (Dussoix and Arber, 1962). Moreover, the first sequenced genome was that of an *Escherichia coli* phage:  $\phi$ X174 (Sanger et al., 1977). For some years, the interest for phages was limited to ecological studies and proposals for their use as indicators of fecal pollution (IAWPRC Study Group on Health Related Water Microbiology, 1991; Jofre et al., 2016); in general, bacteriophages have deserved less interest in comparison to their bacterial hosts or to animal viruses.

Nevertheless, the remarkable estimated number of  $10^{31}$  phages on the Earth (Suttle, 2005) is commonly used by researchers to highlight the importance of phages, which are believed to outnumber any other class of biological entity on the planet. Phages have recently re-emerged as powerful regulators of the bacterial populations in natural ecosystems (Fuhrman, 1999). Moreover, because of the appearance of resistances to different antimicrobial agents, their potential use as

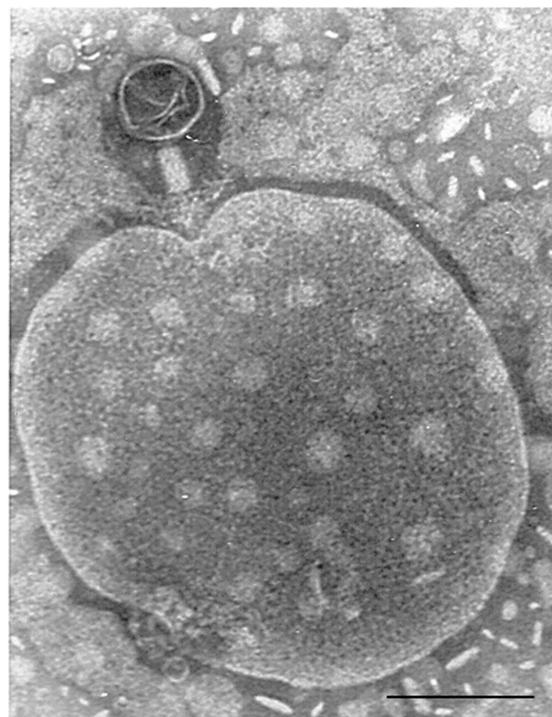
antimicrobials has been revisited (Reardon, 2014). Most significantly, recent metagenomes describe the abundance of viral sequences both outside and inside bacterial cells. This highlights their ubiquity as mobile genetic elements that contribute and affect bacterial evolution, causing the emergence of new bacterial pathogens, mobilizing genes outside the cells, and other functions. Interest in metagenomes includes the study of human microbiomes, where phages again appear as extremely abundant and diverse elements. Researchers are only now starting to suspect that phages actively contribute to the homeostasis of the bacterial flora (De Paepe et al., 2014). Because many studies focus on the role of the human symbiotic microbiota in our wellness, phages thus appear as contributing actors that are directly related with human health (Manrique et al., 2016), and therefore the interest in them is rising.

## PHAGES AS A PART OF HUMAN AND ANIMAL MICROBIOTA

Many metagenomic analyses of human microbiomes show the abundance of phages, which is generally greater than that of eukaryotic viruses. This has been shown in metagenomic analysis of lung, vaginal, skin, oral or intestinal microbiota (Breitbart et al., 2003; Colomer-Lluch et al., 2011a; Minot et al., 2011; Oh et al., 2014; Virgin, 2014). More recently, infectious phages have been found in different clinical samples such as ascitic fluid and urine (Brown-Jaque et al., 2016). It was suggested that they could reach the peritoneal cavity after translocation from the intestine (Górski et al., 2006), where they are present (Figure 1) and abundant. They are also present in voided urine (Brown-Jaque et al., 2016), probably coming from the periurethral area. In animals, phages infecting *Bacteroides* were found in serum (Keller and Traub, 1974), confirming their presence in the blood stream. Translocation of phages from blood to mouse fetal tissues has also been demonstrated in pregnant mice (Srivastava et al., 2004a).

In the light of these results, and as a second level of study, some researchers have analyzed solely the virome fraction of these microbiomes. To do this, they have devised methods that allow discrimination of the viral fraction, while discarding bacterial and free DNA. Those studies have yielded some surprising results; many viral particles in fact carry sequences identified as bacterial DNA. Shared genetic content is observed when analyzing the phage and bacterial DNA fractions of the same sample (Breitbart et al., 2003; Minot et al., 2011; Colombo et al., 2016; Howe et al., 2016), including sequences belonging to CRISPR-Cas systems (Dutilh et al., 2014).

CRISPR-Cas systems constitute an immune system that protect bacteria against bacteriophages and foreign DNA (Mojica and Rodriguez-Valera, 2016), that has later been applied for genome engineering in bacteria and eukaryotes. The different activity of the CRISPR-Cas systems influences the allowance of bacterial cells to foreign DNA or their immunity to phage infection, and this can shape the evolution of human microbiomes. Besides the use of CRISPR-Cas systems in genome engineering, the analysis of CRISPR sequences from raw metagenomic data has revealed unidentified phages, as crAssphage phage, that is claimed to be



**FIGURE 1 | Bacteriophage of *Myoviridae* morphology isolated from a fecal sample, attached to an unidentified particle. Bar 100 nm.**

present in the majority of human fecal microbiomes, although it has never been isolated (Dutilh et al., 2014).

## PHAGES AS MOBILE GENETIC ELEMENTS

Transduction, the process by which the DNA is mobilized between cell by a virus or viral vector was reported the last century (Zinder and Lederberg, 1952), although the rates of this mobilization has never been well defined. For this reason, the detection of an important proportion of bacterial DNA in phage particles observed in metagenomic analysis was indeed a surprise, and it initially prompted the belief that the methods for segregating phage and bacterial particles were not accurate enough, and either bacterial or free DNA contaminated the phage samples. However, the protocols have been optimized allowing specific extraction of packaged DNA. Another suspicion is that the bioinformatic analysis failed to identify phage DNA sequences correctly and they were mistaken for bacterial DNA. Nevertheless, subsequent repetitions and more accurate approaches have shown that despite some of these problems occurring, a relevant fraction of the virome is actually mobilizing bacterial DNA. This has led to the suspicion that bacterial cells use the numerous capsid genes that they possess, probably inherited from ancient prophage remnants, to build protein capsids that pack and spread their DNA content (Asadulghani et al., 2009; Lang et al., 2012; Penades et al., 2015).

The fact that phage capsids can mobilize bacterial DNA has multiple consequences, such as, for example, the fact that they can mobilize and transduce virulence genes (O'Brien et al., 1984; Griffiths et al., 2000; Allué-Guardia et al., 2011; Penades et al., 2015), antibiotic resistances (Muniesa et al., 2004; Colomer-Lluch et al., 2011b; Ross and Topp, 2015; Haaber et al., 2016) or genes related to fitness (Lindell et al., 2004; Müller et al., 2013) to new bacterial hosts. This causes horizontal genetic exchange and leads to the evolution of bacterial populations.

## PHAGES AS REGULATORS OF POPULATIONS

Bacterial populations can change and evolve through acquisition of new genes transferred by phages, but also by predation and lysis caused by phages. Experimental evidence from chemostats and observations of phages/hosts in open systems has shown that for some bacterial species, populations of phages and hosts oscillate over time, following a “Red Queen/kill-the-winner” dynamics,” which describes prey–predator variations (Rodríguez-Brito et al., 2010; Jover et al., 2013; Lim et al., 2015). However, phage–host dynamics can change in accordance with the homogeneity and structure of the environment, and also depending on the conditions that facilitate phage–cell encounters (De Paepe et al., 2014).

Changes or a total replacement of the microbiome by a fecal transplant in diseases without a well-defined etiological agent, such as inflammatory bowel diseases (Crohn's disease or ulcerative colitis), can result in different disease outcomes (Loh and Blaut, 2012; Moayyedi et al., 2015). Comparison of the viromes of individuals suffering from Crohn's disease and healthy relatives revealed differences in composition and variability (Pérez-Brocal et al., 2013; Wagner et al., 2013). Whether changes in the phagome of human biomes is a cause or a consequence of dysbiosis in such diseases has not yet been established. Considering that the phagome could influence bacterial populations, two options are plausible: changes in bacteria could cause variations in the distribution of phage groups; or changes in the phagome could be responsible of dysbacteriosis (Norman et al., 2015; Pérez-Brocal et al., 2015).

Similarly, phages have been detected in the metagenomes of sputum of patients suffering cystic fibrosis (Willner et al., 2009); and both the phage diversity and relative abundances were reported to be different from those of non-cystic fibrosis patients. It is hard, however, to conclude from these results what the cause of these differences is. Some variations in bacteria are caused by phages and those variations could be harmful to the patients. For example, mucoid isolates of *Pseudomonas fluorescens* are more virulent than their non-mucoid isogenic variants. This mucoid overproduction is a virulence factor contributing to more persistent infections in cystic fibrosis patients (Scanlan and Buckling, 2012). This phenotypic characteristic is favorably selected in the presence of phages, because it confers protection against phage infection. Accordingly, the mucoid isolates became resistant to the phages with the corresponding detrimental consequence for the patients (Scanlan and Buckling, 2012).

A different example of the regulation of human bacterial populations by phages is observed when we look at the competition between *Streptococcus pneumoniae* and *Staphylococcus aureus*. The former produces hydrogen peroxide; an agent that induces the bacterial SOS response and can induce temperate prophages. Meanwhile, the vast majority of *S. aureus* strains carry prophages that could be induced in the presence of the concentrations of H<sub>2</sub>O<sub>2</sub> produced by *S. pneumoniae*. *S. pneumoniae* prophages, in turn, are not induced at these concentrations. The result is that *S. pneumoniae* prevails by killing *S. aureus* lysogenic strains via induction of prophages that cause the subsequent lysis of the cell (Selva et al., 2009).

Yet another example of how bacteriophages can impact the dynamics of bacterial populations has been observed in *Enterococcus faecalis* V583. This strain produces a composite phage ΦV1/7, derived from two distinct chromosomally encoded prophage elements. Prophage ΦV1 produces the capsids, while prophage ΦV7 is in charge of infection of susceptible hosts and V583 can produce infectious ΦV1/7. The induction of ΦV1/7 is highly enhanced by the availability of free amino acids in the medium. The strain producing ΦV1/7 has an advantage over other *E. faecalis* strains in the intestine, because these are lysed by ΦV1/7, while V583 is resistant to superinfection, enhancing the success of *E. faecalis* V583 during competitive growth (Duerkop et al., 2012).

## INTERACTIONS WITH THE IMMUNE SYSTEM

It is not clear whether phages can easily be detected by the immune system, or whether they interact with it. Because the size of phage particles is usually bigger than eukaryotic viruses, activation of the immune system might occur as for other viruses. The desire to use phages to treat bacterial infections has led to explorations of the responses that phages might cause within the human immune system.

Very soon after the discovery of phages, it was observed that antibodies against bacteriophages in humans or animals were produced (Jerne, 1952, 1956); and it is easy to generate phage antisera by immunization of humans or animals with phage lysates (Puig et al., 2001; Gorski et al., 2012; Bacon et al., 2017). The sera of non-immunized individuals (humans or animals) present antibodies against phages, although at low levels; the so-called “natural antibodies.” For instance, antibodies against T4 phages are naturally present in human serum (Dabrowska et al., 2014) presumably as a consequence of the confirmed constant presence of phages in human biomes (Górski et al., 2006; Brown-Jaque et al., 2016). However, the origin of natural antibodies, generally of IgM class, with broad cross reactivity and low affinity, is not clear in the majority of cases.

The innate immune system, particularly by the components of the reticuloendothelial system (RES), could be a mechanism for removing phages that are circulating in the human body (Gorski et al., 2012). Certainly, this system was credited with the rapid removal of administered wild-type phage λ from the circulatory system in humans (Geier et al., 1973). Moreover, different phage



$\lambda$  mutants could induce different host responses. When using certain phage  $\lambda$  mutants that were capable of circumventing the RES immune response, these mutants prevailed for longer periods in the blood stream than the wild-type phage (Merrill et al., 1996).

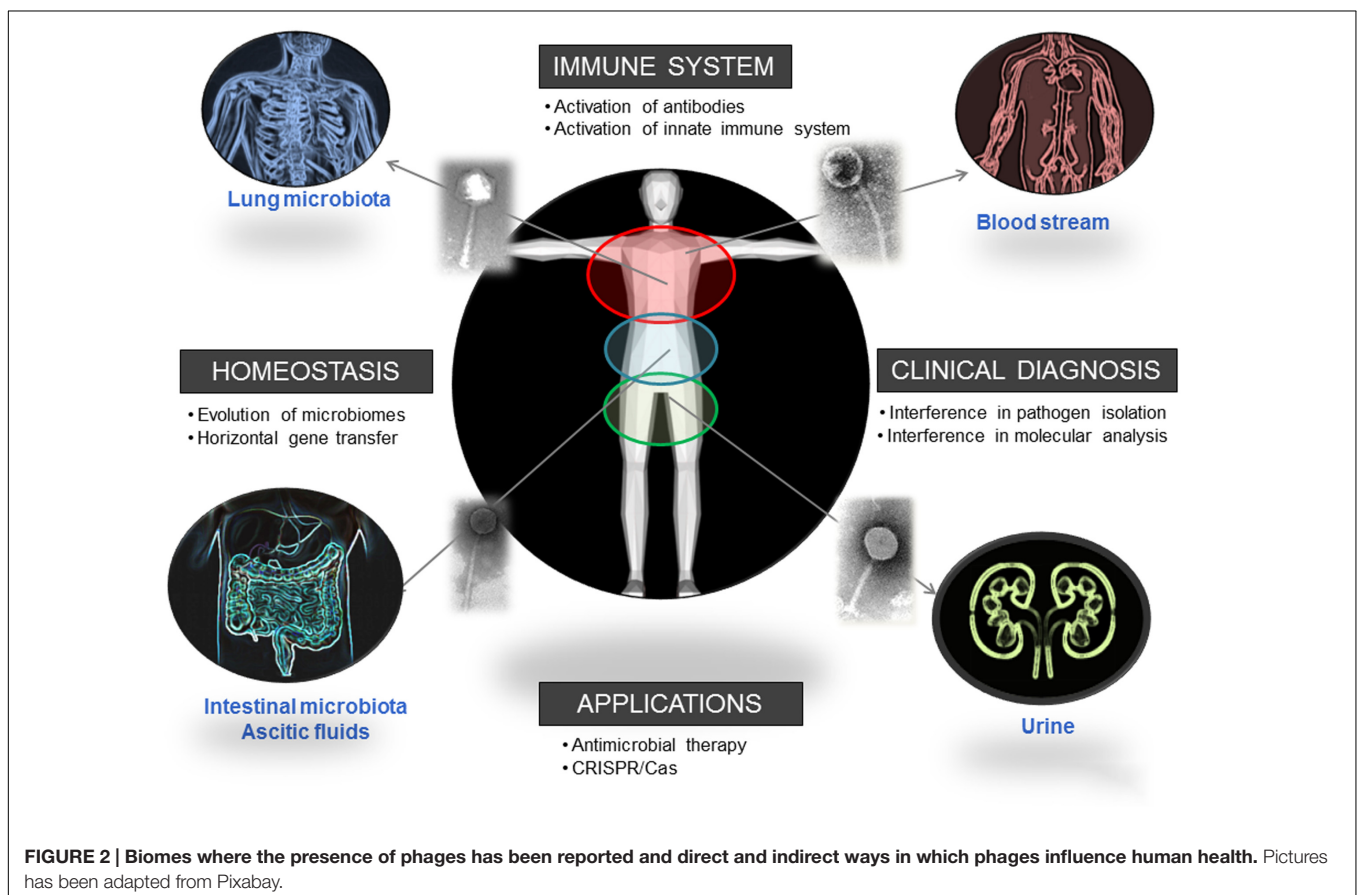
Data on anti-phage cellular responses are very scarce in comparison with data on phage–humoral responses. The only study we are aware of, evaluated the cellular response to MS2 phage that was intradermally administrated in guinea pigs. The presence of the phage produced erythema and induration, that are signs of cell-mediated immunity (Langbeheim et al., 1978). In contrast, another study showed that the permanence of phages in blood is the same when comparing immunocompetent mice or those deficient in T-cells, indicating no specific role of T-cell response in phage inactivation (Srivastava et al., 2004b).

When administered together with the host bacteria, some studies showed that phages seem to stimulate bacterial phagocytosis, and this is attributed to certain “opsonization” of the bacterial cells by phages. In addition, phages can remain active and infective when adsorbed onto the bacteria on intake by granulocytes. Therefore, some authors have suggested that during phagocytosis, phages continue lysing the phagocytosed bacteria, helping the activity of phagocytic cells. This process is limited in time and phages are no longer active after the completion of phagocytosis (Gorski et al., 2012). Despite these descriptions, there is no definitive evidence that phages activate phagocytosis

by themselves, and some years ago, a contrary outcome was reported (Kantoch et al., 1958). In those studies, when used at very high doses ( $10^{10}$ /ml), phages inhibited phagocytosis of their host bacteria, and this inactivation was observed using either infectious or heat-inactivated phages (Kantoch et al., 1958). Inhibition was greater when using antibody-treated phages, and therefore the authors suggested that the immunocomplexes phage–antibody would be inactivating factors particularly active (Kantoch et al., 1958). Moreover, purified phages have anti-inflammatory effects via suppression of ROS (reactive oxygen species) production and inhibition of NF- $\kappa$ B activity, affecting the production of cytokines [for a review, see (Gorski et al., 2012)]. Despite this evidence, it should be borne in mind that many experiments have been conducted with phage lysates, which on many occasions could contain remnants of bacteria lysed by the phages (e.g., lipopolysaccharide) or perhaps fragments of the host bacterial cell wall adhered to the phage tails. This makes it extremely difficult to determine the components truly responsible for the modulation of the immune response.

## INTERFERENCE WITH CLINICAL DIAGNOSES

Assuming the relative occurrence and distribution of phages throughout the human body described above, coincident with





the location of their bacterial hosts, and highly plausible translocation of phage particles to other areas of the body, some reports indicate that the neglected presence of phages in human samples could have an important influence by interfering in clinical practice (Brown-Jaque et al., 2016).

It has been shown that the presence of phages could interfere with many protocols intended to isolate bacteria by enrichment broth, since the phages in the sample destroy the bacterial cells during the enrichment procedure (Muniesa et al., 2005; Quiros et al., 2015). Phages might also interfere in clinical settings during bacterial isolation. As indicated above, there is evidence of a lack of, or reduced, bacterial isolation in clinical samples (ascitic fluid and urine) carrying a high titer of phages, because the lytic activity of the phages disturbs the isolation of the target bacteria (Brown-Jaque et al., 2016).

In addition, some results obtained using molecular methods when targeting some virulence genes present in pathogenic bacteria could be confusing. This is because some genes are located in temperate phages and DNA extraction methods do not distinguish between bacterial and phage DNA. This is the case for phages encoding the Shiga toxin gene, which can be detected in the absence of Shiga toxin-producing bacteria (Martínez-Castillo and Muniesa, 2014). Detection of certain bacterial groups by 16SrDNA qPCR or by genomic sequencing in mixed samples might also be confusing if the sample contains phages and the DNA in the phages is actually what is amplified in the absence of intact bacterial cells. This might be an explanation of the mismatch between the high number of gene copies of 16SrDNA obtained by qPCR amplification and the lack of bacterial isolation observed sometimes (Esparcia et al., 2011). Among others, one hypothesis could be that the positive results were due to amplification of bacterial DNA within phage particles or of bacterial DNA released in the sample after phage-mediated lysis.

## USE OF PHAGES AGAINST HUMAN BACTERIAL PATHOGENS

The problems of fighting antibiotic resistance in bacteria are continually increasing and severely undermine our capacity to control bacterial infectious diseases. After the increased incidence of bacterial resistance to antibiotics over recent decades, phages have surfaced again as alternative or complementary therapies to control bacterial infections (Fischetti et al., 2006; Doyle and Erickson, 2012; Hertwig et al., 2013;

Reardon, 2014; Schmelcher and Loessner, 2014; Górski et al., 2016).

## CONCLUDING REMARKS

Phages, the most abundant entities on the planet, are also present in human biomes. This presence is known and recognized, but sometimes neglected; and it has a strong influence on the distribution and dynamics of different bacterial populations. Considering the influence of these populations in human health, as their reported ability to improve digestive health, it is clear that phages can be directly related with human well-being (Figure 2). The influence of phages in different mechanisms of our immune system suggests a long-term relationship that we are just starting to elucidate. Moreover, considering our interest in isolating and identifying bacterial pathogens, the presence of phages could certainly interfere with that analysis if it is not considered. A One Health multidisciplinary approach, not restricted to academic or clinical settings and not limited either to microbiological studies, is advisable to evaluate the real extent of and the role played by the phagome in human bodies.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This study was supported by the Generalitat de Catalunya (2009SGR1043), the Centre de Referència en Biotecnologia (XeRBA), the Sira Carrasco Foundation Grant and by the Ministerio de Ciencia e Innovación, Instituto de Salud Carlos III, cofinanced by the European Development Regional Fund, A Way To Achieve Europe, ERDF; the Fondo de Investigación Sanitaria (grant PI16/00158) and project MINECO AGL2016-75536-P (AEI/FEDER, EU).

## ACKNOWLEDGMENT

Authors thank Prof. P. Coll and Prof. J. Jofre for useful comments on the manuscript.

## REFERENCES

- Allué-Guardia, A., García-Aljaro, C., and Muniesa, M. (2011). Bacteriophage-encoding cytolethal distending toxin type V gene induced from nonclinical *Escherichia coli* isolates. *Infect. Immun.* 79, 3262–3272. doi: 10.1128/IAI.05071-11
- Asadulghani, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., Iguchi, A., et al. (2009). The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.* 5:e1000408. doi: 10.1371/journal.ppat.1000408
- Bacon, E. J., Richmond, S. J., Wood, D. J., Stirling, P., Bevan, B. J., and Chalmers, W. S. (2017). Serological detection of phage infection in *Chlamydia psittaci* recovered from ducks. *Vet. Rec.* 119, 618–620.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., et al. (2003). Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185, 6220–6223. doi: 10.1128/JB.185.20.6220-6223.2003
- Brown-Jaque, M., Muniesa, M., and Navarro, F. (2016). Bacteriophages in clinical samples can interfere with microbiological diagnostic tools. *Sci. Rep.* 6:33000. doi: 10.1038/srep33000

- Colombo, S., Arioli, S., Guglielmetti, S., Lunelli, F., and Mora, D. (2016). Virome-associated antibiotic-resistance genes in an experimental aquaculture facility. *FEMS Microbiol. Ecol.* 92:fiw003. doi: 10.1093/femsec/fiw003
- Colomer-Lluch, M., Imamovic, L., Jofre, J., and Muniesa, M. (2011a). Bacteriophages carrying antibiotic resistance genes in fecal waste from cattle, pigs, and poultry. *Antimicrob. Agents Chemother.* 55, 4908–4911. doi: 10.1128/AAC.00535-11
- Colomer-Lluch, M., Jofre, J., and Muniesa, M. (2011b). Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS ONE* 6:e17549. doi: 10.1371/journal.pone.0017549
- Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G. Z., Boling, L., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5, 4498. doi: 10.1038/ncomms5498
- Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature* 192, 1227–1232. doi: 10.1038/1921227a0
- Dabrowska, K., Miernikiewicz, P., Piotrowicz, A., Hodyra, K., Owczarek, B., Lecion, D., et al. (2014). Immunogenicity studies of proteins forming the T4 phage head surface. *J. Virol.* 88, 12551–12557. doi: 10.1128/JVI.02043-14
- De Paepe, M., Leclerc, M., Tinsley, C. R., and Petit, M.-A. (2014). Bacteriophages: an underestimated role in human and animal health? *Front. Cell. Infect. Microbiol.* 4:39. doi: 10.3389/fcimb.2014.00039
- D'Herelle, F. (1917). sur un microbe invisible antagoniste des bacilles disenterique. *C. R. Acad. Sci. Ser. D* 165, 373–375.
- Doyle, M. P., and Erickson, M. C. (2012). Opportunities for mitigating pathogen contamination during on-farm food production. *Int. J. Food Microbiol.* 152, 54–74. doi: 10.1016/j.jfoodmicro.2011.02.037
- Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. M., and Hooper, L. V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17621–17626. doi: 10.1073/pnas.1206136109
- Dussoix, D., and Arber, W. (1962). Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda. *J. Mol. Biol.* 5, 37–49. doi: 10.1016/S0022-2836(62)80059-X
- Esparcia, O., Montemayor, M., Ginovart, G., Pomar, V., Soriano, G., Pericas, R., et al. (2011). Diagnostic accuracy of a 16S ribosomal DNA gene-based molecular technique (RT-PCR, microarray, and sequencing) for bacterial meningitis, early-onset neonatal sepsis, and spontaneous bacterial peritonitis. *Diagn. Microbiol. Infect. Dis.* 69, 153–160. doi: 10.1016/j.diagmicrobio.2010.10.022
- Fischetti, V. A., Nelson, D., and Schuch, R. (2006). Reinventing phage therapy: are the parts greater than the sum? *Nat. Biotechnol.* 24, 1508–1511. doi: 10.1038/nbt1206-1508
- Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548. doi: 10.1038/21119
- Geier, M. R., Trigg, M. E., and Merrill, C. R. (1973). Fate of bacteriophage lambda in non-immune germ-free mice. *Nature* 246, 221–223. doi: 10.1038/246221a0
- Gorski, A., Miedzybrodzki, R., Borysowski, J., Dabrowska, K., Wierzbicki, P., Ohams, M., et al. (2012). Chapter 2—“Phage as a modulator of immune responses: practical implications for phage therapy,” in *Advances in Virus Research. Bacteriophages Part B*, eds M. Lobočka and W. Szybalski (San Diego, CA: Academic Press), 41–72.
- Górski, A., Międzybrodzki, R., Weber-Dąbrowska, B., Fortuna, W., Letkiewicz, S., Rogóż, P., et al. (2016). Phage therapy: combating infections with potential for evolving from merely a treatment for complications to targeting diseases. *Front. Microbiol.* 7:1515. doi: 10.3389/fmicb.2016.01515
- Górski, A., Wazna, E., Dabrowska, B.-W., Dabrowska, K., Switała-Jeleń, K., and Miedzybrodzki, R. (2006). Bacteriophage translocation. *FEMS Immunol. Med. Microbiol.* 46, 313–319. doi: 10.1111/j.1574-695X.2006.00044.x
- Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. (2000). “Gene transfer in bacteria and their viruses,” in *An Introduction to Genetic Analysis*, 7th Edn, ed. W. Freeman (New York, NY: W. H. Freeman).
- Haaber, J., Leisner, J. J., Cohn, M. T., Catalan-Moreno, A., Nielsen, J. B., Westh, H., et al. (2016). Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nat. Commun.* 7:13333. doi: 10.1038/ncomms13333
- Hertwig, S., Hammerl, J. A., Appel, B., and Alter, T. (2013). Post-harvest application of lytic bacteriophages for biocontrol of foodborne pathogens and spoilage bacteria. *Berl. Münch. Tierärztl. Wochenschr.* 126, 357–369.
- Howe, A., Ringus, D. L., Williams, R. J., Choo, Z.-N., Greenwald, S. M., Owens, S. M., et al. (2016). Divergent responses of viral and bacterial communities in the gut microbiome to dietary disturbances in mice. *ISME J.* 10, 1217–1227. doi: 10.1038/ismej.2015.183
- IAWPRC Study Group on Health Related Water Microbiology (1991). Bacteriophages as model viruses in water quality control. *Water Res* 25, 529–545. doi: 10.1016/0043-1354(91)90126-B
- Jerne, N. K. (1952). Bacteriophage inactivation by antiphage serum diluted in distilled water. *Nature* 169, 117–118. doi: 10.1038/169117b0
- Jerne, N. K. (1956). The presence in normal serum of specific antibody against bacteriophage T4 and its increase during the earliest stages of immunization. *J. Immunol.* 76, 209–216.
- Jofre, J., Lucena, F., Blanch, A., and Muniesa, M. (2016). Coliphages as model organisms in the characterization and management of water resources. *Water* 8, 199. doi: 10.3390/w8050199
- Jover, L. F., Cortez, M. H., and Weitz, J. S. (2013). Mechanisms of multi-strain coexistence in host-phage systems with nested infection networks. *J. Theor. Biol.* 332, 65–77. doi: 10.1016/j.jtbi.2013.04.011
- Kantoch, M., Skurski, A., and Wiecek, Z. (1958). In vitro blockade of bacterial phagocytosis of leukocytes by means of bacterial viruses. *Schweiz. Z. Pathol. Bacteriol.* 21, 1106–1119. doi: 10.1159/000160571
- Keller, R., and Traub, N. (1974). The characterization of *Bacteroides fragilis* bacteriophage recovered from animal sera: observations on the nature of *Bacteroides* phage carrier cultures. *J. Gen. Virol.* 24, 179–189. doi: 10.1099/0022-1317-24-1-179
- Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472–482. doi: 10.1038/nrmicro2802
- Langbeheim, H., Teitelbaum, D., and Arnon, R. (1978). Cellular immune response toward MS-2 phage and a synthetic fragment of its coat protein. *Cell. Immunol.* 38, 193–197. doi: 10.1016/0008-8749(78)90046-1
- Lim, E. S., Zhou, Y., Zhao, G., Bauer, I. K., Droit, L., Ndao, I. M., et al. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* 21, 1228–1234. doi: 10.1038/nm.3950
- Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11013–11018. doi: 10.1073/pnas.0401526101
- Loh, G., and Blaut, M. (2012). Role of commensal gut bacteria in inflammatory bowel diseases. *Gut Microbes* 3, 544–555. doi: 10.4161/gmic.22156
- Manrique, P., Bolduc, B., Walk, S. T., van der Oost, J., de Vos, W. M., and Young, M. J. (2016). Healthy human gut phageome. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10400–10405. doi: 10.1073/pnas.1601060113
- Martínez-Castillo, A., and Muniesa, M. (2014). Implications of free Shiga toxin-converting bacteriophages occurring outside bacteria for the evolution and the detection of Shiga toxin-producing *Escherichia coli*. *Front. Cell. Infect. Microbiol.* 4:46. doi: 10.3389/fcimb.2014.00046
- Merrill, C. R., Biswas, B., Carlton, R., Jensen, N. C., Creed, G. J., Zullo, S., et al. (1996). Long-circulating bacteriophage as antibacterial agents. *Proc. Natl. Acad. Sci. U.S.A.* 93, 3188–3192. doi: 10.1073/pnas.93.8.3188
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., et al. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625. doi: 10.1101/gr.122705.111
- Moayyedi, P., Surette, M. G., Kim, P. T., Libertucci, J., Wolfe, M., Onischi, C., et al. (2015). Fecal Microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology* 149, 102–109.e6. doi: 10.1053/j.gastro.2015.04.001
- Mojica, F. J. M., and Rodríguez-Valera, F. (2016). The discovery of CRISPR in archaea and bacteria. *FEBS J.* 283, 3162–3169. doi: 10.1111/febs.13766
- Müller, M. G., Ing, J. Y., Cheng, M. K.-W., Flitter, B. A., and Moe, G. R. (2013). Identification of a phage-encoded Ig-binding protein from invasive *Neisseria meningitidis*. *J. Immunol.* 191, 3287–3296. doi: 10.4049/jimmunol.1301153
- Muniesa, M., Blanch, A. R., Lucena, F., and Jofre, J. (2005). Bacteriophages may bias outcome of bacterial enrichment cultures. *Appl. Environ. Microbiol.* 71, 4269–4275. doi: 10.1128/AEM.71.8.4269-4275.2005

- Muniesa, M., García, A., Miró, E., Mirelis, B., Prats, G., Jofre, J., et al. (2004). Bacteriophages and diffusion of beta-lactamase genes. *Emerg. Infect. Dis.* 10, 1134–1137. doi: 10.3201/eid1006.030472
- Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., et al. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 160, 447–460. doi: 10.1016/j.cell.2015.01.002
- O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W., and Formal, S. B. (1984). Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* 226, 694–696. doi: 10.1126/science.6387911
- Oh, J., Byrd, A. L., Deming, C., Conlan, S., Kong, H. H., and Segre, J. A. (2014). Biogeography and individuality shape function in the human skin metagenome. *Nature* 514, 59–64. doi: 10.1038/nature13786
- Penades, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R. P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* 23, 171–178. doi: 10.1016/j.mib.2014.11.019
- Pérez-Brocá, V., García-López, R., Nos, P., Beltrán, B., Moret, I., and Moya, A. (2015). Metagenomic analysis of Crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential interactions and biomarkers. *Inflamm. Bowel Dis.* 21, 2515–2532. doi: 10.1097/MIB.0000000000000549
- Pérez-Brocá, V., García-López, R., Vázquez-Castellanos, J. F., Nos, P., Beltrán, B., Latorre, A., et al. (2013). Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clin. Transl. Gastroenterol.* 4, e36. doi: 10.1038/ctg.2013.9
- Puig, A., Araujo, R., Jofre, J., and Frias-Lopez, J. (2001). Identification of cell wall proteins of *Bacteroides fragilis* to which bacteriophage B40-8 binds specifically. *Microbiology* 147, 281–288. doi: 10.1099/00221287-147-2-281
- Quiros, P., Martínez-Castillo, A., and Muniesa, M. (2015). Improving detection of Shiga toxin-producing *Escherichia coli* by molecular methods by reducing the interference of free Shiga toxin-encoding bacteriophages. *Appl. Environ. Microbiol.* 81, 415–421. doi: 10.1128/AEM.02941-14
- Reardon, S. (2014). Phage therapy gets revitalized. *Nature* 510, 15–16. doi: 10.1038/510015a
- Rodríguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., et al. (2010). Viral and microbial community dynamics in four aquatic environments. *ISME J.* 4, 739–751. doi: 10.1038/ismej.2010.1
- Ross, J., and Topp, E. (2015). Abundance of antibiotic resistance genes in bacteriophage following soil fertilization with dairy manure or municipal biosolids, and evidence for potential transduction. *Appl. Environ. Microbiol.* 81, 7905–7913. doi: 10.1128/AEM.02363-15
- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., et al. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265, 687–695. doi: 10.1038/265687a0
- Scanlan, P. D., and Buckling, A. (2012). Co-evolution with lytic phage selects for the mucoid phenotype of *Pseudomonas fluorescens* SBW25. *ISME J.* 6, 1148–1158. doi: 10.1038/ismej.2011.174
- Schmelcher, M., and Loessner, M. J. (2014). Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage* 4:e28137. doi: 10.4161/bact.28137
- Selva, L., Viana, D., Regev-Yochay, G., Trzcinski, K., Corpa, J. M., Lasa, I., et al. (2009). Killing niche competitors by remote-control bacteriophage induction. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1234–1238. doi: 10.1073/pnas.0809600106
- Srivastava, A. S., Chauhan, D. P., and Carrier, E. (2004a). In utero detection of T7 phage after systemic administration to pregnant mice. *Biotechniques* 37, 81–83.
- Srivastava, A. S., Kaido, T., and Carrier, E. (2004b). Immunological factors that affect the in vivo fate of T7 phage in the mouse. *J. Virol. Methods* 115, 99–104.
- Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160
- Twort, F. W. (1915). An investigation on the nature of ultra-microscopic viruses. *Lancet* 186, 1241–1243. doi: 10.1016/S0140-6736(01)20383-3
- Virgin, H. W. (2014). The virome in mammalian physiology and disease. *Cell* 157, 142–150. doi: 10.1016/j.cell.2014.02.032
- Wagner, J., Maksimovic, J., Farries, G., Sim, W. H., Bishop, R. F., Cameron, D. J., et al. (2013). Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing. *Inflamm. Bowel Dis.* 19, 1598–1608. doi: 10.1097/MIB.0b013e318292477c
- Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F. E., Silva, J., et al. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE* 4:e7370. doi: 10.1371/journal.pone.0007370
- Zinder, N. D., and Lederberg, J. (1952). Genetic exchange in *Salmonella*. *J. Bacteriol.* 64, 679–699.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Navarro and Muniesa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Replication of Staphylococcal Resistance Plasmids

Stephen M. Kwong<sup>1\*</sup>, Joshua P. Ramsay<sup>2</sup>, Slade O. Jensen<sup>3</sup> and Neville Firth<sup>1</sup>

<sup>1</sup> School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia, <sup>2</sup> School of Biomedical Sciences, Curtin University, Perth, WA, Australia, <sup>3</sup> Antimicrobial Resistance and Mobile Elements Group, Ingham Institute for Applied Medical Research, Sydney, NSW, Australia

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Sultan Zainal Abidin University,  
Malaysia

### Reviewed by:

Elisabeth Grohmann,  
Beuth University of Applied Sciences  
Berlin, Germany  
Gloria Del Solar,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### \*Correspondence:

Stephen M. Kwong  
stephen.kwong@sydney.edu.au

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 September 2017

**Accepted:** 06 November 2017

**Published:** 23 November 2017

### Citation:

Kwong SM, Ramsay JP, Jensen SO  
and Firth N (2017) Replication  
of Staphylococcal Resistance  
Plasmids. *Front. Microbiol.* 8:2279.  
doi: 10.3389/fmicb.2017.02279

The currently widespread and increasing prevalence of resistant bacterial pathogens is a significant medical problem. In clinical strains of staphylococci, the genetic determinants that confer resistance to antimicrobial agents are often located on mobile elements, such as plasmids. Many of these resistance plasmids are capable of horizontal transmission to other bacteria in their surroundings, allowing extraordinarily rapid adaptation of bacterial populations. Once the resistance plasmids have been spread, they are often perpetually maintained in the new host, even in the absence of selective pressure. Plasmid persistence is accomplished by plasmid-encoded genetic systems that ensure efficient replication and segregational stability during cell division. Staphylococcal plasmids utilize proteins of evolutionarily diverse families to initiate replication from the plasmid origin of replication. Several distinctive plasmid copy number control mechanisms have been studied in detail and these appear conserved within plasmid classes. The initiators utilize various strategies and serve a multifunctional role in (i) recognition and processing of the cognate replication origin to an initiation active form and (ii) recruitment of host-encoded replication proteins that facilitate replisome assembly. Understanding the detailed molecular mechanisms that underpin plasmid replication may lead to novel approaches that could be used to reverse or slow the development of resistance.

**Keywords:** staphylococci, multiresistance plasmid, plasmid replication, replication initiation protein, plasmid copy number control, antisense RNA

## INTRODUCTION

Plasmids are accessory extra-chromosomal genetic elements that provide bacteria with various adaptive qualities that have contributed to their success in diverse environmental niches. Over the last seven decades, the use of antimicrobial compounds in medical, veterinary, and agricultural practices has provided strong evolutionary selection for the acquisition of pre-existing and newly evolved antimicrobial resistance genes for bacterial survival. Plasmids have been instrumental in the dissemination of these resistance genes, and the rapid evolution of multiply resistant strains of *Staphylococcus aureus* in hospitals throughout the world provides an exemplar of this process.

*Staphylococcus aureus* is commonly normal flora of healthy individuals but is capable of causing serious life-threatening conditions, predominantly in debilitated individuals or patients undergoing surgical procedures. They have long been a primary cause of nosocomial infections and are notorious for the propensity to develop resistance to multiple antimicrobial agents.



Particularly concerning has been the emergence of multidrug-resistant community-associated strains of *S. aureus* capable of causing highly virulent infections in healthy populations (Nimmo, 2012; Planet, 2017). In staphylococci, resistance genes are primarily associated with mobile genetic elements such as plasmids, genomic islands, and transposons (Lyon and Skurray, 1987; Schwarz et al., 2014). Many of these elements are capable of horizontal transfer between bacterial cells by conjugation, mobilization, and/or phage-mediated mechanisms, thus promoting the spread of resistance genes. In *S. aureus*, conjugative plasmids play a central role in enabling intercellular DNA transfer of both conjugative and mobilizable plasmids, which are each capable of accruing multiple resistance and/or virulence genes. It has recently been demonstrated that many of the large staphylococcal multiresistance plasmids, that were previously thought to be non-mobilizable, can in fact be mobilized through the carriage of *oriT* mimic sequences (O'Brien et al., 2015; Ramsay et al., 2016; Ramsay and Firth, 2017). The transfer of these types of plasmids can facilitate multidrug resistance evolution in a single step. Thus, both conjugative and mobilizable plasmid classes are important adaptive tools that have had a major impact on the evolution of antimicrobial resistance.

In the event of plasmid horizontal transfer, a single-stranded copy of the plasmid is transmitted from the donor cell to a recipient cell where it is re-circularized and replicated into a double-stranded form. If the transferred plasmid is capable of efficient autonomous replication in the recipient and becomes established, the plasmid and its resistance genes are rarely lost. Most small plasmids counteract loss during cell division by replicating at high copy numbers. In this situation, there are many plasmid copies that are randomly distributed to daughter cells as the cytoplasmic contents of the parent cell are shared. It is uneconomical for larger plasmids to use this strategy due to the metabolic and genetic loads imposed upon the host cell that would encumber fitness. Large plasmids instead encode various segregational stability systems, often including active partitioning, post-segregational killing, and multimer resolution, that work together to maintain extremely high inherent stability, enabling them to replicate with low copy numbers, presumably to minimize the burden on the host. Crucial to plasmid survival, in terms of both segregational stability and fitness cost, is replication, which can be initiated when needed during cell division and turned off once the ideal plasmid copy number is established after division. This essential capacity to sense and adjust plasmid quantity leads to a defined average copy number in any given host. Plasmids display considerable diversity in their replication systems, the components used to initiate replication and the mechanisms by which replication is controlled (del Solar et al., 1998; Brantl, 2014). Comprehensive reviews on the replication of circular bacterial plasmids (del Solar et al., 1998) and of rolling-circle plasmids (Khan, 1997; Ruiz-Masó et al., 2015) are available in the literature. This article reviews our current understanding of the replication initiation mechanisms of staphylococcal plasmids (last reviewed by Novick, 1989), the systems they use to control plasmid copy number, and

includes an updated view of the distribution of *S. aureus* plasmid types.

## CLASSIFICATION OF STAPHYLOCOCCAL PLASMIDS

Staphylococcal plasmids range from just over 1 kb to greater than 60 kb in size. In general, the smaller plasmids (between 1 and 8 kb) are cryptic or encode a single resistance determinant and replicate via a rolling-circle replication (RCR) mechanism that is hallmarked by the production of single-stranded intermediates during replication. Many similarities can be drawn between the replication of RCR plasmids and some classes of bacteriophages (e.g.,  $\Phi$ X174), with the major difference being the control of replication frequency. Bacteriophage often propagates their genomes in bursts without consideration to the host's survival, whereas plasmids are host-dependent and thus control their replication to be synergetic with that of the host. Staphylococcal RCR plasmids (also known as class I staphylococcal plasmids) can be further sub-grouped into plasmid families based on the evolutionary relationships of their essential replication initiation proteins. Four main subclasses of RCR plasmids have been found in staphylococci and are represented by the prototypes pT181, pC194, pE194, and pSN2, which are distinguished by the type of replication initiation gene that they carry (Novick, 1989; Firth and Skurray, 2006). Some RCR plasmids have been extensively studied including closely related tetracycline resistance plasmid pT181 and chloramphenicol resistance plasmid pC221 of the pT181 family (Thomas et al., 1995; Khan, 1997). Characterized members of the pC194 family include pC194 itself, conferring resistance to chloramphenicol (Gruss et al., 1987), and the aminoglycoside resistance plasmid pUB110 (Maciag et al., 1988). The best studied plasmid of the pE194 family is tetracycline resistance plasmid pMV158, originally isolated from streptococci but shown to be capable of stable replication in a wide range of bacterial species (Meijer et al., 1995). pSN2 family members have been studied to a lesser extent in regard to their replication and control mechanisms.

Staphylococcal plasmids greater than 8 kb in size typically utilize a theta ( $\theta$ )-type replication mechanism and have historically been divided into two main classes depending upon their conjugative ability (Novick, 1989; Firth and Skurray, 2006). Non-conjugative theta-replicating plasmids include the well-known  $\beta$ -lactamase/heavy-metal resistance plasmids, pSK1-family multiresistance plasmids, and pSK639-family plasmids (Firth and Skurray, 2006). The conjugative plasmids have traditionally included closely related multiresistance plasmids exemplified by pSK41 (Liu et al., 2013) and pGO1 (Caryl and O'Neill, 2009); however, two new distinct families have recently been described (Ramsay et al., 2016). In general, the conjugative plasmids are larger than the non-conjugative plasmids due to the carriage of extensive gene arrays that encode a type IV secretion system (T4SS), a large multiprotein pore complex through which single-stranded DNA can be transferred to the recipient cell, a nicking relaxase enzyme and its DNA substrate (*oriT*), relaxase

accessory proteins, and a coupling protein that provides the basis of recognition between the relaxase and the mating pore. RCR plasmids of the pT181, pC194, and pE194 families and some members of the theta-replicating pSK639-family are known to carry only the relaxase unit (*pre* or *mob* genes and *oriT*), which enables horizontal transfer via mobilization. Plasmid mobilization can occur if these plasmids are able to exploit the mating pore provided by a suitable co-resident conjugative plasmid (or other conjugative element).

## PLASMID INCOMPATIBILITY

Non-identical plasmids that share nearly identical replication/maintenance components (DNA, RNA, and/or proteins) display plasmid incompatibility. That is, they are unable to be maintained efficiently through continued rounds of cell division in the absence of plasmid selection. Incompatibility is caused by the inability of the *trans*-acting replication or maintenance components to distinguish “self” from “non-self,” and has historically been used as an indication of plasmid relatedness. Staphylococcal plasmids have been placed into at least 15 incompatibility groups (Ruby and Novick, 1975; Iordanescu and Surdeanu, 1980; Novick, 1989; Udo and Grubb, 1991). Ten groups corresponded to RCR plasmids with the remainder being larger, and hence, probably theta-replicating. It was noted earlier that very closely related RCR plasmids were in different incompatibility groups and quite dissimilar theta-replicating plasmids were often in the same incompatibility group (Novick, 1989). This indicated that while incompatibility tests can yield biologically relevant information regarding the ability of plasmids to coexist stably, they do not necessarily indicate relatedness on a whole, particularly in regard to their resistance or other phenotypes. The recombinatory systems (described below) that have shaped the evolution of both RCR and theta-replicating plasmids provide an explanation for this phenomenon.

## EVOLUTION OF RESISTANCE PLASMIDS

As the nucleotide sequences of RCR plasmids became available, it was apparent that they were composed of interchangeable modules or gene cassettes (Gruss and Ehrlich, 1989; Novick, 1989). Genetic features, such as resistance genes, mobilization systems, and lagging-strand replication origins were not conserved among plasmids of the same incompatibility group or replicon type. The gene cassette junctions were noted to be abrupt with the level of sequence identity dropping from near perfect to no homology across a single pair of nucleotides (Gruss and Ehrlich, 1989; Novick, 1989). The production of ssDNA in the RCR mechanism is critical to this type of cassette exchange due to the greatly increased capacity for homologous and illegitimate recombination events (Niaudet et al., 1984; Janni  re and Ehrlich, 1987). Thus, a cassette can insert or replace another cassette if the appropriate

flanking sequences are present in the target plasmid. An example of cassette dissemination by this mechanism is evident with the multidrug resistance locus *qacC* in members of the pC194 family where the gene is located between conserved elements required for leading and lagging strand replication (Leelaporn et al., 1995; Wassenaar et al., 2016). The composition and arrangement of gene cassettes and the associated flanking regions are shuffled by occurrences such as plasmid co-integration and aberrant replication events to generate new combinations (Ballester et al., 1989; Gruss and Ehrlich, 1989).

pT181 was discovered to have defined recombination sites, termed  $RS_A$  and  $RS_B$ , that promote the formation of plasmid co-integrates by site-specific recombination (Novick et al., 1984).  $RS_B$  is a conserved sequence found in the pT181 lagging strand replication origin *pala* (aka *ssoA*), which is broadly carried by diverse staphylococcal RCR plasmids (Novick et al., 1989). Plasmids bearing lagging strand replication origin types *ssoU*, *ssoT*, and *ssoW* also possess the conserved  $RS_B$  sequence (Kramer et al., 1999).  $RS_A$  is found in the region upstream of *pre*, the protein product (Pre) of which is essential for recombination at this site (Gennaro et al., 1987). Later it was discovered that Pre is a member of the pMV158 Mob protein family, which was essential for the mobilization of this plasmid (Priebe and Lacks, 1989). Therefore, it is now understood that the primary role of pT181 Pre is in mobilization and its role in recombination is a by-product of its nicking activity at  $RS_A$ , which functions as the pT181 *oriT*. Pre-mediated nicking of  $RS_A$  generates an efficient substrate for co-integrate formation between two plasmids (Priebe and Lacks, 1989).

Theta-replicating plasmids are known to carry transposons but often also contain one or more assemblies of resistance gene clusters inserted into a conserved plasmid backbone (Firth and Skurray, 2006). These regions often correspond to IS257-flanked co-integrated copies of small plasmids. The genetic arrangement of clusters resembles that of composite transposons but probably have not inserted as such. Rather the clusters are likely the result of IS257 transposition by a non-resolved replicative mechanism leading to co-integration of two plasmids and resulting in directly repeated IS257 copies at each junction (Needham et al., 1995; Leelaporn et al., 1996; Firth and Skurray, 1998). Often the resulting co-integrand is then fine tuned by intra-molecular transposition of IS257, which promotes sequence deletions of flanking DNA (e.g., the replication region of co-integrated RCR plasmids; Berg et al., 1998), and sequences within the terminal inverted repeats of IS257 can modulate the expression of adjoining genes by generating hybrid promoters (Leelaporn et al., 1994; Simpson et al., 2000; P  rez-Roth et al., 2010). Thus, theta-replicating plasmids accrue resistance genes through the incorporation of smaller resistance plasmids and subsequent deletion of problematic or unnecessary sequences (Firth and Skurray, 1998). In addition to influencing the evolution of theta-replicating multiresistance plasmids, IS257-mediated RCR plasmid integration events also appear to have played a role in the evolution of staphylococcal genomic islands, such as SCC*mec* (Stewart et al., 1994; Firth and Skurray, 2006).

## PLASMID REPLICATION REGIONS

Staphylococcal plasmids carry a 1- to 2-kb region that contains genetic information for autonomous replication and its control. The essential components include (i) an origin of replication (*dso* in RCR plasmids or *ori* in theta-replicating plasmids), (ii) a replication control element (antisense RNA and/or protein), and (iii) a gene encoding the replication initiation protein, Rep. In RCR plasmids, the double-stranded origin, *dso*, contains a sequence-specific binding site for the Rep protein and a short, partially palindromic sequence. In pT181, Rep binding to the *dso* alters the conformation of this palindromic sequence to a cruciform structure that is efficiently nicked by Rep (Koepsel and Khan, 1986; Noirot et al., 1990). In most RCR plasmids, the Rep binding site and the nick site are adjacent, and in other plasmids (e.g., pMV158), the two sites are separated by a short distance of up to 100 bp. The *dso* is often positioned upstream of the *rep* coding region, except in pT181 family plasmids, where the *dso* is found within the *rep* coding region (Figure 1). In RCR plasmids, an additional element, the single-stranded origin, *sso*, is required for efficient lagging strand synthesis using the leading strand as template. Therefore, the *sso* is only functional on the leading strand and its orientation cannot be reversed (Gruss et al., 1987). The *sso* displays more variety in its position relative to the *rep* gene (Figure 1) and is often separated from it by various gene cassettes indicating that its position is flexible. *sso* sequences are known to limit the host range of RCR plasmids and currently five classes are known to exist: *ssoA*, *ssoU*, *ssoT*, *ssoW*, and *ssoL* (Ruiz-Masó et al., 2015). The ~160 nt *ssoA* carried by pT181 family plasmids restricts their stable replication to staphylococcal species (Gruss et al., 1987) and *ssoA* sequences carried by other plasmid families also appear host-restricted (Kramer et al., 1998). In a single-stranded state, the *ssoA* is capable of forming a large secondary structure that is recognized and primed by RNA polymerase (RNAP) (Kramer et al., 1997). In addition to DNA polymerase III (PolC), DNA polymerase I (PolA) is required during elongation and termination of lagging strand synthesis (Diaz et al., 1994; Kramer et al., 1997). Members of the pC194 and pE194 families have been shown to carry a widely recognized RNAP-dependent *ssoU* permitting them to replicate in all *firmicutes* tested (Boe et al., 1989; Kramer et al., 1995; Lorenzo-Díaz and Espinosa, 2009). *ssoW* of lactococcal plasmid pWVO1 appears to be capable of both RNAP-dependent priming and primosome-dependent priming via a *ssoW*-located primosome assembly site; however, lagging strand replication of this plasmid is restricted to lactococcal species (Seegers et al., 1995).

In theta-replicating plasmids, the position of the origin of replication (*ori*) is variable. In pSK41, the *ori* has been shown to be contained centrally within *rep*, whereas in pSK639 and pCH91 family members, the predicted location of *ori* is upstream and downstream of *rep*, respectively (Figure 1). Although *dso* and *ori* both contain one or more binding sites for the replication initiator, they are functionally quite different. Cleavage of the parental DNA strands does not occur during theta-replication and instead a short region of the *ori* (or immediately next to

it) is melted upon Rep binding to form an open single-stranded initiation complex.

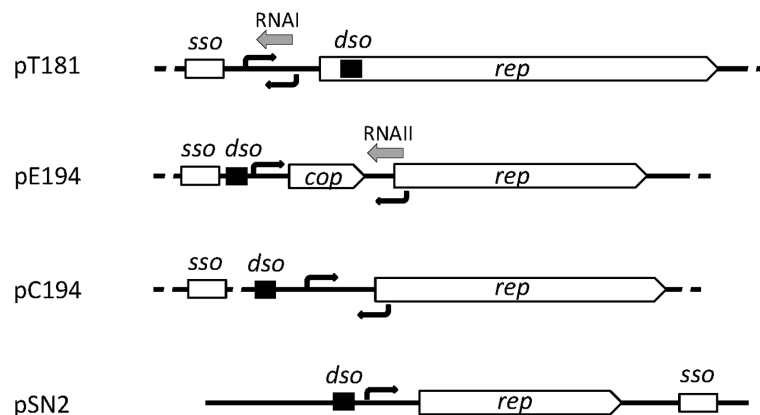
## REPLICATION INITIATORS

To date, staphylococcal plasmids have been found to encode one or more of seven distinct types of replication initiation protein, defined by the conserved domains Rep\_trans (pfam02486; pT181), Rep\_1 (pfam01446; pC194), Rep\_2 (pfam01719; pE194), and RepL (pfam05732; pSN2) for the RCR plasmids, and Rep\_3 (pfam01051; pSK639), RepA\_N (pfam06970; pSK41), and PriCT\_1 (pfam08708; pCH91) for the theta-replicating plasmids. The domain organizations for representatives of each protein family are illustrated in Figure 2. All of the initiators contain DNA-binding domains (DBD) for specific binding to their cognate *dso* or *ori*. In some cases, oligomerization domains (OD) have been identified. The RCR initiators are all expected to have topoisomerase activity essential for *dso* cleavage at initiation of replication and for cleavage and ligation during the termination of leading strand synthesis. Conserved catalytic tyrosine residues needed for these activities are also indicated where known (Figure 2).

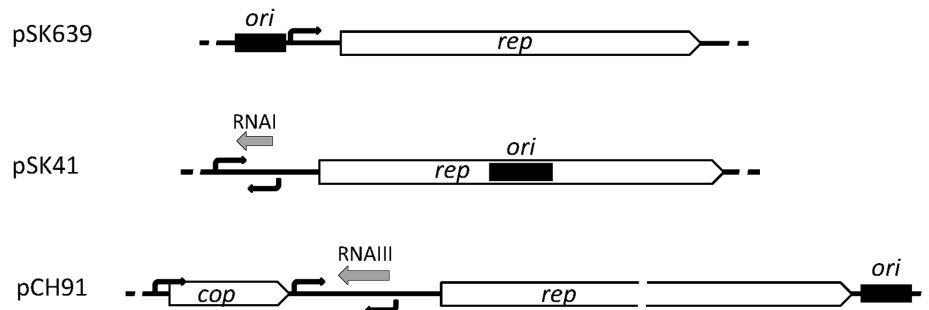
The large number of completely sequenced staphylococcal plasmids now available has allowed us to present an updated view of the distribution of replication initiator genes that are typical of each plasmid family. Due to its clinical significance, plasmids from *S. aureus* clearly dominate the databases. It is currently not known whether distribution of the plasmid classes presented here would vary markedly in other species of the genus. Our analysis of the replication genes of 278 completely sequenced non-identical *S. aureus* plasmids is presented in Figure 3 (Supplementary Data Sheet 1). Slightly more than half of the plasmids (57%) are predicted to replicate via a theta-type mechanism with the remainder using an RCR mechanism (Figure 3). As noted previously (Novick, 1989), the RCR plasmids appear restricted in size with > 90% less than 5 kb. Of the RCR plasmids, 45% were expected to utilize an initiator containing the Rep\_1 conserved domain, 28% Rep\_trans, 22% RepL, and only 3% Rep\_2, with two novel plasmids that could not be grouped. Interestingly, there also appears to be a trend between RCR initiator type and plasmid size where RepL < Rep\_1 < Rep\_trans although there are also numerous exceptions to this trend (Figure 3).

Of 160 *S. aureus* theta-replicating plasmids, those using RepA\_N initiators are clearly the most common, being encoded by approximately 80% (Figure 3). Rep\_3 domain initiators are encoded by approximately 20% and PriCT\_1 by about 6% of sequenced plasmids. A total of 11% of theta-replicating plasmids contained two different potential replication initiation genes of these types. In some cases, there is evidence that one of the replicons has been inactivated through genetic alterations (mutations/truncations). In plasmids with both *repA\_N*- and *rep\_3*-type genes, there are examples where either type appears inactivated. In other cases, there is no obvious defect in either replication region. In regard to the conjugative plasmids, use of RepA\_N-type initiators is even more pronounced being used by 90% (18 out of 20), with only two cases of a PriCT\_1-type

## Rolling-Circle Replicating Plasmids



## Theta-Replicating Plasmids



**FIGURE 1 |** Genetic organization of staphylococcal plasmid replication regions. A representative of each of the seven plasmid initiator types are illustrated and include the *rep* gene for the replication initiation protein, *cop* gene encoding a *rep* repressor, origin of replication (*dso* or *ori*, black rectangles), predicted or determined promoters (black arrows), antisense RNAs (gray arrows), and single-stranded origin of replication (*sso*).

replicon. Interestingly, there are no Rep<sub>3</sub>-type plasmids that are predicted to be conjugative.

## ROLLING-CIRCLE REPLICATION

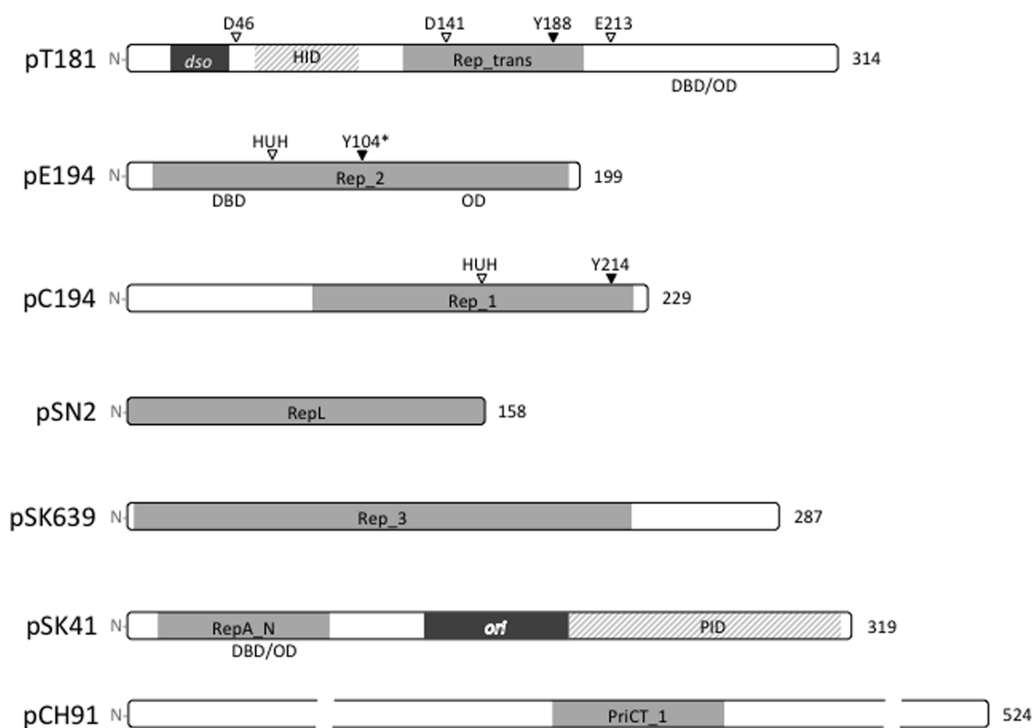
### Rep<sub>trans</sub> Plasmids – the pT181 Family

pT181 family plasmids are widespread in the staphylococci and frequently carry tetracycline or chloramphenicol resistance genes. The two best studied members include pT181 and pC221, and their respective initiators, RepC and RepD, share 82% overall sequence identity. Sequence identity (90%) increases across the common Rep<sub>trans</sub> domains, which contain the essential catalytically active tyrosine residues (RepC Y188; RepD Y191). The major region of sequence divergence between the initiators lies in the C-terminal regions, which, for both proteins, has been shown to bind to their cognate *dsos* (Koepsel and Khan, 1986; Thomas et al., 1990). These variable regions have been shown to contain a six amino acid determinant that governs *dso*-binding specificity and limits cross reactivity between initiators and

origins of different Rep<sub>trans</sub> plasmids (Dempsey et al., 1992; Wang et al., 1992). As such, multiple incompatibility groups exist in the pT181 plasmid family potentially allowing them to stably coexist in the same host cell.

In pT181 family plasmids, the Rep protein binds as a dimer to the *dso* bending the DNA into a cruciform structure that exposes the nick site in its single-stranded form (Noirot et al., 1990). The topoisomerase activity of the initiator results in nicking at the *dso* and formation of a phosphodiester bond between the reactive tyrosine hydroxyl group and the newly created 5' phosphate of the nicked strand (Thomas et al., 1990). The Rep protein also aids recruitment of PcrA helicase, which is essential to plasmid replication and unwinds and separates the leading parental strand from the lagging template strand (Slatter et al., 2009). Synthesis is initiated from the Rep-generated 3' hydroxyl group and is known to require DNA polymerase III (Majumder and Novick, 1988). Synthesis of the leading strand continues full circle and 10–12 nt past the original nick site (Rasooly and Novick, 1993). The covalently attached Rep dimer then catalyzes a second series of strand-transfer





**FIGURE 2 |** Replication initiation proteins of staphylococcal plasmids. The initiation proteins of prototypical plasmids are represented by bars that indicate the relative sizes of the proteins. Conserved domains within each plasmid replication initiator type are shaded gray. The HUH and DDE metal ion-binding motifs and catalytic tyrosine residues (Y) are labeled where known. \*The Y104 of pE194 was predicted by generating an amino acid sequence alignment with the pMV158 RepB protein. In pT181 and pSK41 plasmid families, the origins of replication (*dso* and *ori*) are located within the *rep* coding regions and derived protein sequences in these regions are dispensable for replication. The pT181 RepC–PcrA interaction domain (HID) and the pSK41 Rep–DnaG primase interaction domain (PID) are indicated. Where known, the positions of DNA-binding domains (DBD) and oligomerization domains (OD) are also shown.



**FIGURE 3 |** Distribution of replicon types among sequenced *Staphylococcus aureus* plasmids. The graph at the top indicates plasmid sizes for 278 plasmids in the databases. The conserved domains for each replication initiation protein are indicated by a black/gray bar below the graph. Some plasmids encode more than one Rep type. Plasmids predicted to be non-conjugative are represented by black bars and plasmids predicted to be conjugative are represented by gray bars.

reactions that terminates leading strand replication, resulting in a double-stranded plasmid containing a newly synthesized leading strand, a re-circularized single-stranded displaced strand and a Rep dimer (RepC/RepC\*) that remains attached to the 10–12 nt adduct rendering the protein inactive for replication (Rasooly et al., 1994). This inactivation mechanism prevents reuse of the initiator for further rounds of replication and is a fundamental prerequisite for pT181 plasmid copy number control.

Crystallography and 3D modeling of Rep\_trans initiators suggests that the dimeric structure resembles a horseshoe with a basic channel sitting atop the structure (containing the six amino acid specificity determinant) that could accommodate dsDNA (Carr et al., 2016). The catalytic residues, required for DNA nicking and ligation, including the Tyr active site, are positioned on the inner face. The divalent metal ion-binding site that is essential for topoisomerase activity is coordinated by three distally located residues DDE (**Figure 2**) brought together

within antiparallel  $\beta$ -sheets (Carr et al., 2016). Residues affecting PcrA helicase interaction were mapped to the open end of the horseshoe (Carr et al., 2016).

## Rep\_1 and Rep\_2 Plasmids – the pC194 and pE194 Families

Although possessing distinguishable conserved domains, the Rep\_1 and Rep\_2 families of plasmid replication initiators are united by the HUH metal ion-binding motif and a catalytic Tyr residue which are both necessary for topoisomerase activity (Chandler et al., 2013). The HUH superfamily also includes bacteriophage replication proteins, conjugative DNA relaxases and transposases. Members of the HUH superfamily may have either one or two active tyrosine residues (Chandler et al., 2013). Plasmids in both families exhibit a broad host range with some members having been shown to replicate efficiently in a wide range of Gram-positive and Gram-negative bacteria and even eukaryotic cells (Weisblum et al., 1979; Goursoot et al., 1982; del Solar et al., 1987; Coffey et al., 1994; Aleshin et al., 1999).

The pMV158 initiator RepB (210 residues) has been studied in detail and is one of only a small number of Rep proteins with structure determined by crystallography (Boer et al., 2009). Unlike the dimeric state of pT181 initiators, pMV158 RepB is instead shown to form a hexameric ring structure that may encircle the DNA and increase processivity of leading strand synthesis (Ruiz-Masó et al., 2004; Boer et al., 2009). The N-terminal region of RepB contains both the DNA binding and topoisomerase activities of the protein, while the C-terminal domain is required for hexameric oligomerization (Boer et al., 2016). The pMV158 *dso* contains a set of three directly repeated sequences (*bind*) located 84 bp downstream of the nick site, which is found in the loop of a palindromic secondary structure (Puyet et al., 1988). It has been shown that, at least *in vitro*, RepB can bind to DNA fragments containing either the nick site or the *bind* locus, although the affinity of the protein for the latter is much higher (Ruiz-Masó et al., 2007). On the other hand, RepB catalytic activity requires a single-stranded DNA substrate, and the requirement for plasmid supercoiling indicates that the nick site is presented in a single-stranded conformation only when in this topological state (del Solar et al., 1987; Moscoso et al., 1995). In contrast to the Rep\_trans type initiators, the active Tyr residue in pMV158 RepB does not appear to form a stable covalent phosphodiester bond with the 5'-end of the cleaved strand (Moscoso et al., 1995). However, a more labile RepB-DNA covalent adduct was observed after rapid treatment of cleavage reactions with SDS and proteinase K (Ruiz-Masó et al., 2016), indicating that RepB inactivation after one round of leading strand synthesis might still occur through a similar mechanism.

The pC194 and pUB110 replication initiation proteins, RepA and RepU, are monomers in solution but bind to their *dsos* cooperatively in pairs (Noirot-Gros et al., 1994; Müller et al., 1995). At high concentrations, approximately six RepU monomers coat the *dso* region and this complex is able to extend upon the adjacent *repU* promoter leading to *repU* transcriptional silencing (Müller et al., 1995). The pC194 RepA protein has three catalytic residues, Tyr214, Glu142, and Glu210 that are essential

for nicking and closing activities of the protein but that do not affect *dso* binding and these residues are universally conserved in all plasmids of the pC194 family (Noirot-Gros et al., 1994). The proposed roles of the essential residues are as follows: (i) nucleophilic attack at the nick site during initiation by Tyr214; (ii) hydrolysis of the regenerated nick site at the leading strand termination step promoted by Glu210 (which would act as a general base catalyst); and (iii) metal ion coordination involving Glu142 (Noirot-Gros et al., 1994). The equivalent residues in pUB110 RepU are Tyr241, Glu163, and Glu237 (Noirot-Gros et al., 1994; **Figure 3**). The 37 kDa RepU protein has two forms that show slightly different molecular weights in polyacrylamide gels and it has been suggested that the larger form could be attached to a short oligo (like pT181) or modified in another way that renders it inactive for replication (Müller et al., 1995). Although the precise mechanism used to prevent Rep recycling has not yet been elucidated, such a mechanism would be a prerequisite for effective copy number control (Gros et al., 1989). Interestingly, it has been reported that the Rep proteins of several RCR plasmids, including pC194, can also act as mobilization relaxases in the presence of the integrative conjugative element ICEBs1 (Lee et al., 2012).

## RepL Plasmids – the pSN2 Family

pSN2 family members are among the smallest RCR plasmids in staphylococci (**Figure 3**). Many pSN2 family plasmids are cryptic or carry a single resistance determinant, with erythromycin resistance being the most common. No members of this family are known to carry mobilization functions. The pSN2 *dso* and *ssmA* can be readily identified based on similarity to other RCR plasmids (Khan and Novick, 1982; Novick, 1989; Khan, 1997) and the *ssmA* was shown to be required for efficient lagging strand synthesis by an RNAP-dependent priming mechanism (Dempsey et al., 1995). Notably the pSN2 family RepL proteins contain a helix-turn-helix (HTH) DNA domain and are predicted to be about 18 kDa, whereas other RCR initiators are considerably larger (Novick, 1989; Catchpole and Dyke, 1992). At least two incompatibility groups have been shown to exist in the pSN2 family (Oliveira et al., 1993) with sequence variation most often occurring in the central part of the protein (residues 85–95), indicating the location of a possible DNA-binding specificity determinant. At this time, no copy number control mechanism has been identified although many pSN2 family members carry a short open reading frame (that is only sometimes annotated) encoding a small, basic protein (45–60 aa) that could potentially play a regulatory role.

## THETA-MODE REPLICATION

### RepA\_N Plasmids

Plasmid replication initiators containing the conserved RepA\_N domain are frequently used by theta-replicating *S. aureus* multiresistance plasmids, both conjugative and non-conjugative, and are widespread in many other coagulase-negative staphylococcal species, including plasmids from *S. epidermidis*, *S. haemolyticus*, *S. saprophyticus*, *S. xylois*,

and *S. warneri*. The RepA\_N-type initiators are also broadly distributed among other large plasmids of the low G+C Gram-positive firmicutes such as enterococci, lactobacilli, lactococci, and bacilli (Firth et al., 2000; Weaver et al., 2009).

The best studied RepA\_N plasmid is the conjugative multiresistance plasmid pSK41, which confers resistance to the aminoglycosides kanamycin, tobramycin, gentamycin, and neomycin, as well as to bleomycin and antiseptics and disinfectants (Berg et al., 1998; Liu et al., 2013). The pSK41 Rep protein (319 aa) has been divided into three functional domains. The N-terminal 120 residues (NTD) contains the conserved RepA\_N domain and mediates *ori* specific binding at four Rep boxes found centrally within the *rep* gene (Kwong et al., 2004; Liu et al., 2012). The NTD also contains sequences necessary for oligomerization of the protein (Schumacher et al., 2014). The central domain (121–199 aa), encoded by DNA sequences corresponding to the *ori*, functions as a linker and proteins carrying an in-frame deletion of the central domain can rescue replication of Rep-defective plasmids containing a functional *ori* (Liu et al., 2012). The Rep C-terminal domain (CTD; 200–319) is essential for replication and displays considerably high sequence conservation, but only in plasmids from the same genera, suggesting that it may perform a host-specific function (Weaver et al., 2009). Recently, it was shown that the pSK41 Rep CTD interacts directly with the *S. aureus* DnaG primase (Schumacher et al., 2014).

Crystallographic analysis of RepA\_N proteins indicated that the Rep NTD readily forms tetramers and contains a winged HTH that allows interaction with both the major and the minor grooves of Rep box DNA, inducing a bend of approximately 30° (Schumacher et al., 2014). This was consistent with the intrinsic bend associated with A+T-rich tracts and substitution of this Rep box sequence to a G+C-rich tract reduced pSK41 Rep binding by 10-fold (Schumacher et al., 2014). The Rep CTD was shown to form a compact structure composed of five helices and on its own was monomeric. Interestingly, the pSK41 Rep structures of the NTD and CTD were found to display structural similarity to the primosomal protein DnaD, suggesting a common evolutionary origin (Schumacher et al., 2014). Structural similarity between the *B. subtilis* DnaD and DnaB proteins that was not evident in pairwise sequence alignments has also been described (Marston et al., 2010). In *B. subtilis*, DnaD and DnaB play a central role in both DnaA-mediated initiation of replication at *oriC* and also in restart of stalled replication forks by primosome assembly (Bruand et al., 1995; Smits et al., 2011).

Multiple incompatibility groups are likely to exist within the staphylococcal RepA\_N plasmids since pSK41 and pSK1, which encode divergent RepA\_N proteins, are known to be compatible (Firth et al., 2000). A novel chimeric replication initiation gene was identified in the high-level mupirocin resistance plasmid pPR9 from Spain (Pérez-Roth et al., 2010). pPR9 displays high-level nucleotide sequence homology to pSK41 throughout the backbone including replication, maintenance, and transfer regions, except for a ~500-bp region corresponding to the pSK41 RepA\_N domain and replication origin (Pérez-Roth et al., 2010). The pPR9 Rep NTD was instead found to share homology to a putative phage replication protein while retaining > 97%

amino acid sequence identity to the pSK41 Rep CTD and 99% nucleotide sequence identity in the upstream *rep* control region. Instead of the winged-HTH present in the RepA\_N domain, the pPR9 Rep NTD contains a putative HTH belonging to pfam13730. Construction of a pPR9 mini-replicon showed that the pPR9 *rep* region supported autonomous replication and that it was compatible with a pSK41 mini-replicon (Pérez-Roth et al., 2010). This intriguing modular pPR9 initiator represents a new initiator type that could have evolved to overcome incompatibility barriers. This type of hybrid initiation gene is found in a small number of other plasmids such as pUSA03 from caMRSA strain USA300 (Diep et al., 2006).

## Rep\_3 Plasmids

One fifth of theta-replicating *S. aureus* plasmids were found to carry a replication initiation gene that gives rise to a product containing the conserved Rep\_3 domain (pfam01051). In staphylococci, this type of initiator was first observed on the small (8 kb) *S. epidermidis* trimethoprim resistance plasmid pSK639 (Apisiridej et al., 1997). The pSK639 *rep* gene encodes a protein of 287 residues in length with the conserved Rep\_3 domain spanning the first ~220 residues and a ~50 residue CTD. Like the RepA\_N plasmids, Rep\_3 domain initiators are widely distributed in plasmids from low G+C Gram-positive bacteria including coagulase-positive and -negative staphylococci, enterococci, and lactococci and are distantly related to a large number of iteron-regulated plasmid initiators of Gram-negative bacteria including *Pseudomonas syringae* plasmid pPS10, *Escherichia coli* plasmids F, pSC101, R6K, P1, and broad host range plasmid RK2. The Gram-negative Rep\_3 initiators are variously dependent upon DnaA for their replication and often contain dimerization motifs critical for copy number control (Ingmer et al., 1995; Matsunaga et al., 1997; del Solar et al., 1998; Toukdarian and Helinski, 1998; Das and Chatteraj, 2004; Giraldo and Fernández-Tresguerres, 2004; Swan et al., 2006; Konieczny et al., 2014).

Upstream of the pSK639 *rep* coding region, in the vicinity of the promoter, is a series of five, 22 bp tandemly repeated sequences that most likely represent Rep binding sites and constitute the origin of replication (Apisiridej et al., 1997). The position of these potential Rep binding sites suggests that pSK639 Rep may autoregulate its own transcription (Apisiridej et al., 1997) although this has yet to be demonstrated. Iteron-mediated regulation has been shown to be the primary copy number control mechanism in many of the Rep\_3 domain plasmid replicons from Gram-negative bacteria and could also play a role in pSK639 copy number regulation. This form of regulation relies upon dimerization domains that allow plasmids to pair (handcuff). In these plasmids, Rep proteins are only active for initiation as monomers and at higher Rep concentrations dimerization promotes plasmid pairing and the inhibition of replication initiation.

## PriCT\_1 Plasmids

A smaller number of staphylococcal theta-replicating plasmids (~6%) encode a replication initiator belonging to the broad host range Inc18 family, which includes the enterococcal

conjugative plasmid pAM $\beta$ 1 and streptococcal conjugative resistance plasmids pSM19035 and pIP501. The three plasmids share a high degree of sequence identity (Lioy et al., 2010) and utilize closely related Rep proteins containing the conserved PriCT\_1 domain (Figure 2), which are considerably larger than most plasmid replication initiators (~60 kDa). Staphylococcal plasmids that were detected to contain a PriCT\_1 type replication initiator include pCH91 (17 kb), which encodes a type II toxin-antitoxin system *pemIK* (Bukowski et al., 2013), the exfoliative toxin B plasmid pETB (38 kb; Yamaguchi et al., 2001), pWBG707 (Udo et al., 1992), and *cfr*-carrying conjugative plasmid pSA737 (39 kb; Mendes et al., 2013). The other PriCT\_1 plasmids detected in *S. aureus* were also found to encode a RepA\_N or Rep\_3 initiator gene.

Replication of Inc18 plasmids has been characterized in considerable detail using pAM $\beta$ 1 to study the replication initiation mechanism and pIP501 the copy number control mechanism. These plasmids exhibit a broad host range. In pAM $\beta$ 1 replication, the RepE monomer binds to a single 25-bp sequence in the origin, which is located immediately downstream of the *repE* gene and induces localized melting of a short DNA region (15 nt) found next to the binding site (Le Chatelier et al., 2001). The RepE protein has higher affinity for non-specific single-stranded DNA than for its double-stranded binding site and this activity is believed to play a role in extending *ori* strand opening. Transcription through the *ori* is essential for the replication process, which is independent of DnaA but requires DNA polymerase I (Bruand et al., 1993; Ceglowski et al., 1993; Bruand and Ehrlich, 1998). It has been proposed that the Rep transcript synthesized by RNAP stalls at *ori* when Rep is bound, the transcript is cleaved (by RNAP or Rep) leaving a ~20-nt RNA that acts as the replication primer, which is extended by DNA polymerase I (Le Chatelier et al., 2001). The D-loop structure generated by DNA polymerase I is then an efficient substrate for PriA-mediated primosome assembly that requires the host-encoded replication proteins DnaB, DnaD, and DnaI (Polard et al., 2002). Sequences central to the pAM $\beta$ 1 *ori* (5'-TGCCATTACATTTAT-3') that constitute the RepE binding site (Le Chatelier et al., 2001) and are also found in the minimal *ori* of pIP501 and pSM19035 (Brantl and Behnke, 1992a; Lioy et al., 2010) can be detected in an analogous position downstream of the *rep* genes in staphylococcal plasmids pCH91, pETB, pWBG707, and pSA737, suggesting that they utilize a similar replication initiation mechanism. Furthermore, *cop* and antisense RNA genes similarly positioned to copy number control elements in pIP501 (see below; Brantl and Behnke, 1992b) can also be detected in each of the staphylococcal plasmids.

## COPY NUMBER CONTROL MECHANISMS

Antisense RNA-mediated copy number control is broadly utilized by both RCR and theta-replicating staphylococcal plasmids where regulation of replication has been investigated, including members of the pT181, pE194, pC194, and pSK41 families. Copy number control in the pSN2, pSK639, and pCH91 families have

not yet been studied in detail, however, as described above, members of the latter family appear likely to utilize copy number control systems similar to the Inc18 broad host range conjugative pIP501 employing both a small protein repressor (Cop) and an antisense RNA-mediated attenuation system to regulate Rep expression (reviewed in Brantl, 2014 and Grohmann et al., 2016). Dual-regulation of copy number by both Cop repressor and antisense RNA has also been well established in members of the pE194 family via pMV158 (del Solar and Espinosa, 2000), although the mechanistic details of the systems are distinctive for each family.

## Antisense RNA-Mediated Copy Number Control of Rep<sub>trans</sub> Plasmids

pT181 family plasmids use small, untranslated antisense RNAs to regulate expression of the Rep protein and thereby control replication initiation. In pT181, the 87-nt antisense RNA (RNAI) is counter-transcribed to *repC* and is complementary to the *repC* mRNA untranslated leader region. RNA-RNA interaction between RNAI and the *repC* mRNA causes the formation of a thermodynamically stable secondary structure (stem-loop IV) immediately upstream of the *repC* start codon (Novick et al., 1989). Stem-loop IV is predicted to contain the ribosome binding site, however, the main effect of antisense RNA binding is transcriptional termination at stem-loop IV (which resembles a  $\sigma$ -independent terminator) and the attenuated transcripts are incapable of producing RepC (Novick et al., 1989). In the absence of RNAI, sequences in the 5'-proximal arm of stem-loop IV preferentially pair with another complementary sequence in the *repC* leader, termed the pre-emptor, preventing formation of stem-loop IV and allowing full-length *repC* mRNA to be transcribed (Novick et al., 1989).

## Copy Number Control in Rep<sub>2</sub> Family Plasmids

pE194 family plasmids are predicted to utilize a copy number control system that has been studied extensively in pMV158 and its deletion derivative pLS1. pMV158 encodes two *trans*-acting negative regulators of the replication initiation gene, *repB*, an antisense RNA (RNAII) and a small repressor protein (CopG; del Solar and Espinosa, 1992, 2000). *copG* is found upstream of *repB* and the two genes form an operon. RNAII is a 48-nt transcript that is counter-transcribed from a promoter within the 5'-end of the *repB* coding sequence and is complementary to a region found immediately upstream of an atypical ribosome binding site essential for RepB translation (López-Aguilar et al., 2013). It has been proposed that mRNA-RNAII duplex formation hinders binding of the ribosome to the translation initiation region (López-Aguilar et al., 2013). RNA-RNA interactions initiate through base contacts in the RNAII 5' single-stranded tail, while the RNAII stem-loop appears to only play an auxiliary role in RepB translational repression (López-Aguilar et al., 2015). CopG is dimeric in solution and has a ribbon-helix-helix structure (RHH\_1, pfam01402; Gomis-Rüth et al., 1998). It has been shown that four dimers bind cooperatively to the *copG* promoter leading to transcriptional repression of the *copG*-*repB* transcript



by competitively inhibiting the RNAP-promoter interaction (Hernández-Arriaga et al., 2009). A recent study has indicated ‘crosstalk’ between the pMV158 mobilization and replication systems. MobM was found to bind and repress the RNAII promoter, leading to elevated levels of RepB and an increase in plasmid copy number (Lorenzo-Díaz et al., 2017). Staphylococcal pE194 (Rep\_2) family plasmids including pCPS49, pDLK3, SAP085B each have analogously positioned elements that are predicted to give rise to a CopG-like repressor and a RNAII-like antisense RNA. Thus, the staphylococcal plasmids carrying a Rep\_2 initiator are all expected to utilize similar copy number control system to pMV158. pE194, pCPS49 and SAP085B also carry a *mob/pre* gene distantly related to pMV158 *mobM*. In Rep\_1 family plasmids pC194 and pUB110, antisense RNAs are also predicted to be counter-transcribed in the respective Rep translation initiation regions and have been proposed to directly block translation of the initiator (Alonso and Tailor, 1987; Maciag et al., 1988), although the intricacies of these antisense RNA-mediated control systems have yet to be described in the same detail.

## Antisense RNA-Mediated Copy Number Control of RepA\_N Plasmids

Staphylococcal RepA\_N plasmids that have so far been examined are found to employ closely related antisense RNA-mediated copy number control systems, comparable to the prototype pSK41 (Kwong et al., 2008). It is noteworthy that similar RNA-mediated control systems do not appear to be present in RepA\_N plasmids from other genera. Expression of pSK41 Rep is negatively regulated by a ~83-nt antisense RNA (RNAI) that is counter-transcribed to the *rep* mRNA and is complementary to its leader region in a position ~100 nt upstream of the translation initiation region (Kwong et al., 2004). It has been proposed that binding of RNAI to the Rep mRNA leader induces formation of a stem-loop secondary structure in the *rep* translation initiation region. However, unlike pT181 family plasmids, the antisense RNA induced stem-loop does not attenuate transcription but rather sequesters the ribosome binding site in the stem-loop rendering it inaccessible to the translation machinery (Kwong et al., 2006). Secondary structure probing of pSK41 RNAI revealed the presence of two stem-loops separated by an 8-nt single-stranded spacer and an unstructured 18-nt 5′-tail (Kwong and Firth, 2015). Mutations in either stem-loop significantly reduced RNAI repressor activity but the single-stranded regions could be deleted without affecting RNAI function (Kwong and Firth, 2015), indicating that complete base pairing between the antisense RNA and its target was not required.

## Copy Number Control in PriCT\_1 Plasmids

pIP501 copy number is controlled by two *trans*-acting negative regulators, a small repressor protein, CopR, and an antisense RNA, designated RNAIII (see Figure 1; pCH91). *copR* is found upstream of the initiator gene, *repR*, but the genes are independently transcribed. CopR consists of 92 amino acid residues and contains a conserved HTH domain (HTH\_XRE;

pfam01381) that facilitates operator DNA binding as a dimer (Steinmetzer et al., 1998). CopR does not autoregulate but binds and represses transcription from the *repR* promoter (Brantl, 1994). It has also been demonstrated that CopR-mediated repression of the *rep* promoter effectively increases expression of RNAIII by preventing convergent transcription (Brantl and Wagner, 1997). RNAIII interacts with the leader of the RepR mRNA to induce the formation of a terminator-like structure that results in attenuation of the RepR transcript in a similar manner to that observed in pT181 replication control (Brantl et al., 1993). In *S. aureus* plasmids pCH91, pWBG707, and pSA737 all of the copy number control elements present in pIP501 can be detected even though the predicted replication initiators of the staphylococcal plasmids only share ~30% amino acid sequence identity to pIP501 RepR. These include a small HTH domain protein of the XRE family (Cop), a *rep* promoter that gives rise to a long (~320 nt) leader, an antisense RNA promoter positioned midway through the leader that could give rise to an antisense RNA (Figure 2), and inverted repeats followed by a poly[T] tract just 5′ of the *rep* translation initiation region that appears capable of forming a  $\sigma$ -independent terminator-like structure. The presence of these elements indicates that the staphylococcal PriCT\_1 family plasmids are likely to use an analogous copy number control system to plasmid pIP501.

## HOST-ENCODED PROTEINS IN PLASMID REPLICATION

As we have described above, plasmids encode their own replication components for the initiation of replication, including a replication initiation protein and an origin of replication, and a mechanism that controls the expression/activity of the initiation protein. The interaction between initiator and origin prepares the DNA for replication, either by strand-specific cleavage at *dso* (generating a 3′-OH) or melting of strands at *ori*. Both of these replication mechanisms then depend on helicase enzymes to facilitate further duplex melting. In contrast to RCR plasmids, and as part of the initiation process, theta-replicating plasmids require synthesis of a leading strand replication primer to generate 3′-OH. Once initiated, plasmids then rely on replisomes consisting of host-encoded replication proteins that are normally used for chromosomal replication and repair. In theta-replication of plasmids and the chromosome, the replisome is composed of DNA polymerase III holoenzyme, primase, sliding clamps, helicase, and other accessory factors (Kornberg and Baker, 1992). Fundamental differences between the RC replication mechanism and theta-type replication mechanism (asymmetric vs. semi-conservative) suggest that the replisome components could be quite different. In this section, we discuss some of the host-encoded proteins that are known to play a role in the replication of staphylococcal plasmids.

In most bacteria, DnaA is the essential replication initiator of the chromosomal origin, *oriC*. In many theta-replicating plasmids of *E. coli*, the Rep proteins have been shown to recruit DnaA to their *ori* to assist in the initiation step and often the *oris* possess DnaA boxes homologous to DnaA-binding sites in *oriC*

(del Solar et al., 1998). DnaA has not yet been directly implicated in the replication mechanism of any staphylococcal plasmids and plasmid DnaA boxes have not so far been detected.

## Polymerases

*Escherichia coli* possesses five different DNA polymerases, Pals I, II, III, IV, and V. Pals II, IV, and V are translesion polymerases, Pol III is the core processive polymerase involved in the replisome and Pol I is required in lagging strand theta-replication to remove RNA primers and fill in the gaps of Okazaki fragments (Kornberg and Baker, 1992). Low G+C, Gram-positive bacteria usually possess three DNA polymerase enzymes, PolC, DnaE, and PolA, which are thought to be functionally equivalent to PolIII, PolII, and PolI, respectively. DnaE has been shown to be essential for viability in both *S. aureus* and *B. subtilis* (Dervyn et al., 2001; Inoue et al., 2001). In *B. subtilis*, DnaE was not involved in leading strand synthesis but was essential in lagging strand synthesis for initial extension of RNA primers (Sanders et al., 2010). The role of PolA would likely be in removing RNA primers and joining Okazaki fragments as in *E. coli*. PolA has also been shown to be involved in specific stages of replication of some staphylococcal plasmids including lagging strand synthesis of RCR plasmids (Diaz et al., 1994; Kramer et al., 1997) and initial extension of the leading strand RNA primer in PriCT\_1-family plasmids (Bruand et al., 1993). In both of these stages of plasmid replication, RNAP is involved in generating the replication primer at the *ssu* or by transcription through *ori* as described above.

## Helicases

Bacteria have multiple helicases that have specialized roles (Hall and Matson, 1999). In *E. coli*, DnaB helicase is the primary replicative helicase and is required for replication of the chromosome and theta-replicating plasmids (Kornberg and Baker, 1992). Specialized helicases include the misleadingly named “Rep” helicase protein involved in the replication of some phages (Lane and Denhardt, 1975; Takahashi et al., 1979), and UvrD that engages in DNA repair and replication of some viruses and RCR plasmids (Bierne et al., 1997; Bruand and Ehrlich, 2000). In low G+C Gram-positive bacteria, the DnaB homolog, DnaC, is expected to fulfill the main replicative helicase role.

Rolling-circle replication plasmids of the pT181 family have been shown to require the host-encoded helicase PcrA for replication. Mutations in *S. aureus* *pcrA* led to the accumulation of pT181 initiation complexes indicating that a transition to elongation phase of replication had stalled (Iordanescu and Basheer, 1991). Suppressor mutations that restored replication were mapped to the pT181 Rep protein, suggesting a direct interaction between the two proteins (Iordanescu, 1993). Rep loads the helicase onto the lagging strand of the nicked *dso* and remains engaged with PcrA increasing its processivity and enabling it to displace DNA from a nicked substrate (Soultanas et al., 1999; Chang et al., 2002; Anand and Khan, 2004; Zhang et al., 2007). pT181 Rep displays an interaction with the PcrA helicases of *S. aureus*, *Bacillus anthracis*, *Bacillus cereus*, and *Streptococcus pneumoniae* but fails to stimulate full unwinding activity in the latter (Anand et al., 2004; Ruiz-Masó et al., 2006). Together with previous observations that pT181 can

replicate in bacilli (albeit unstably) but not streptococci, these results indicate that Rep-mediated activation of PcrA is a requirement for efficient replication and lack of interaction is likely to limit plasmid host range. In *B. subtilis* carrying the *pcrA3* mutation, pT181 was incapable of replication; however, pC194 and pE194 plasmids could still replicate at normal copy number. This suggested that either another helicase may be required for leading strand synthesis in pC194 and pE194 or that the *pcrA3* mutation does not effect PcrA interaction with their respective Rep proteins (Petit et al., 1998). Interestingly, in *E. coli* host cells, the PcrA homolog UvrD was found to be essential for pC194 and pE194 replication (Bruand and Ehrlich, 2000).

## Primases

The bacterial primosome is a multiprotein complex containing helicase, primase, and accessory proteins that assist in helicase loading and is required for generating RNA primers on single-stranded DNA. Primosomes are assembled during the initiation of chromosome replication at *oriC* (DnaA-dependent) and in the restart of stalled or collapsed replication forks (PriA-dependent). In Gram-positive bacteria, the replicative helicase (DnaC) is loaded through DnaI with the assistance of DnaB and DnaD. Once DnaC has been loaded it recruits DnaG primase and this primosome complex may then associate with the PolC holoenzyme and other factors to constitute the replisome. pAMβ1 carries a primosome assembly site (*ssiA*) downstream of *ori* that requires PriA, DnaB, DnaD, and DnaI suggesting that the activated *ori* (containing a D-loop) is recognized and targeted by PriA in a process that resembles re-combinational DNA repair (Polard et al., 2002). pSK41 Rep was shown to share structural similarity to DnaD primosomal helicase loader and interact directly with DnaG primase (Schumacher et al., 2014). Thus, it is possible that in pSK41, Rep assists loading of the DnaC helicase, perhaps in combination with DnaB and DnaI, or it may recruit DnaC indirectly through helicase interaction domains conserved in DnaG.

## CONCLUDING REMARKS

The different types of plasmid replication systems described here encompass the diversity of plasmids recognized in staphylococci. Plasmids using each of these systems have been shown to act as vehicles for the carriage of antimicrobial resistance genes. Our view of plasmid diversity in staphylococci is heavily skewed by a historical focus on clinical isolates, and the consequential bias toward *S. aureus* and hence under-representation of coagulase negative species. The extent to which current understanding represents a comprehensive or distorted description of the staphylococcal plasmidome is an open question that awaits far broader sampling of the disparate environments occupied by staphylococci. Advances in sequencing capacity provide an opportunity to address this knowledge gap, while increasing evidence of transmission pathways linking bacteria that impact human health with those in the broader biosphere should provide motivation.

Just as our understanding of plasmid diversity is likely incomplete, the level to which the differing replication systems used by staphylococcal plasmids have been studied varies tremendously. While RCR plasmid replication has been analyzed in considerable detail, the replication systems of theta-replicating plasmids have been largely ignored in comparison, despite the established significance of these plasmids in the expression of resistance and virulence properties. There are several areas where information about these larger staphylococcal plasmids is particularly lacking. This includes how their replication systems interface with the chromosomally encoded replication machinery, and how they integrate and cooperate with other plasmid modules associated with plasmid propagation, such as partitioning, conjugation, and mobilization systems. There seems to be a general view that plasmid biology is well understood, but this is really not the case excepting a handful of model systems. This point has been highlighted by the recent characterization of new types of staphylococcal conjugative plasmids and identification of previously unrecognized mobilization determinants that are widespread on staphylococcal plasmids (O'Brien et al., 2015; Ramsay et al., 2016; Ramsay and Firth, 2017). Given the pivotal role plasmids play in bacterial adaptation, not least in the emergence of staphylococcal

resistance, a renewed emphasis on studies elucidating the properties and mechanisms of plasmids is required if we are to meaningfully appreciate their roles in bacterial evolution and its consequences.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

Research on staphylococcal plasmid biology was supported by National Health and Medical Research Council (Australia) Project Grant APP1081412 to NF, SK, and SJ.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02279/full#supplementary-material>

## REFERENCES

- Aleshin, V. V., Semenova, E. V., Doroshenko, V. G., Jomantas, Y. V., Tarakanov, B. V., and Livshits, V. A. (1999). The broad host range plasmid pLF1311 from *Lactobacillus fermentum* VKM1311. *FEMS Microbiol. Lett.* 178, 47–53. doi: 10.1111/j.1574-6968.1999.tb13758.x
- Alonso, J. C., and Tailor, R. H. (1987). Initiation of plasmid pC194 replication and its control in *Bacillus subtilis*. *Mol. Gen. Genet.* 210, 476–484. doi: 10.1007/BF00327200
- Anand, S. P., and Khan, S. A. (2004). Structure-specific DNA binding and bipolar helicase activities of PcrA. *Nucleic Acids Res.* 32, 3190–3197. doi: 10.1093/nar/gkh641
- Anand, S. P., Mitra, P., Naqvi, A., and Khan, S. A. (2004). *Bacillus anthracis* and *Bacillus cereus* PcrA helicases can support DNA unwinding and *in vitro* rolling-circle replication of plasmid pT181 of *Staphylococcus aureus*. *J. Bacteriol.* 186, 2195–2199. doi: 10.1128/JB.186.7.2195-2199.2004
- Apisiridej, S., Leelaporn, A., Scaramuzzi, C. D., Skurray, R. A., and Firth, N. (1997). Molecular analysis of a mobilizable theta-mode trimethoprim resistance plasmid from coagulase-negative staphylococci. *Plasmid* 38, 13–24. doi: 10.1006/plas.1997.1292
- Ballester, S., Lopez, P., Espinosa, M., Alonso, J. C., and Lacks, S. A. (1989). Plasmid structural instability associated with pC194 replication functions. *J. Bacteriol.* 171, 2271–2277. doi: 10.1128/jb.171.5.2271-2277.1989
- Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A., Leelaporn, A., and Skurray, R. A. (1998). Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. *J. Bacteriol.* 180, 4350–4359.
- Bierne, H., Seigneur, M., Ehrlich, S. D., and Michel, B. (1997). *uvrD* mutations enhance tandem repeat deletion in the *Escherichia coli* chromosome via SOS induction of the RecF recombination pathway. *Mol. Microbiol.* 26, 557–567. doi: 10.1046/j.1365-2958.1997.6011973.x
- Boe, L., Gros, M. F., te Riele, H., Ehrlich, S. D., and Gruss, A. (1989). Replication origins of single-stranded-DNA plasmid pUB110. *J. Bacteriol.* 171, 3366–3372. doi: 10.1128/jb.171.6.3366-3372.1989
- Boer, D. R., Ruiz-Masó, J. A., López-Blanco, J. R., Blanco, A. G., Vives-llàcer, M., Chacón, P., et al. (2009). Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. *EMBO J.* 28, 1666–1678. doi: 10.1038/emboj.2009.125
- Boer, D. R., Ruiz-Masó, J. A., Rueda, M., Petoukhov, M. V., Machón, C., Svergun, D. I., et al. (2016). Conformational plasticity of RepB, the replication initiator protein of promiscuous streptococcal plasmid pMV158. *Sci. Rep.* 6:20915. doi: 10.1038/srep20915
- Brantl, S. (1994). The *copR* gene product of plasmid pIP501 acts as a transcriptional repressor at the essential *repR* promoter. *Mol. Microbiol.* 14, 473–483. doi: 10.1111/j.1365-2958.1994.tb02182.x
- Brantl, S. (2014). Plasmid replication control by antisense RNAs. *Microbiol. Spectr.* 2:PLAS-0001-2013. doi: 10.1128/microbiolspec.PLAS-0001-2013
- Brantl, S., and Behnke, D. (1992a). Characterization of the minimal origin required for replication of the streptococcal plasmid pIP501 in *Bacillus subtilis*. *Mol. Microbiol.* 6, 3501–3510.
- Brantl, S., and Behnke, D. (1992b). Copy number control of the streptococcal plasmid pIP501 occurs at three levels. *Nucleic Acids Res.* 20, 395–400.
- Brantl, S., Birch-Hirschfeld, E., and Behnke, D. (1993). RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. *J. Bacteriol.* 175, 4052–4061. doi: 10.1128/jb.175.13.4052-4061.1993
- Brantl, S., and Wagner, E. G. (1997). Dual function of the *copR* gene product of plasmid pIP501. *J. Bacteriol.* 179, 7016–7024. doi: 10.1128/jb.179.22.7016-7024.1997
- Bruand, C., and Ehrlich, S. D. (1998). Transcription-driven DNA replication of pAM $\delta$ 1 in *Bacillus subtilis*. *Mol. Microbiol.* 30, 135–145. doi: 10.1046/j.1365-2958.1998.01044.x
- Bruand, C., and Ehrlich, S. D. (2000). UvrD-dependent replication of rolling-circle plasmids in *Escherichia coli*. *Mol. Microbiol.* 35, 204–210. doi: 10.1046/j.1365-2958.2000.01700.x
- Bruand, C., Ehrlich, S. D., and Jannière, L. (1995). Primosome assembly site in *Bacillus subtilis*. *EMBO J.* 14, 2642–2650.
- Bruand, C., Le Chatelier, E., Ehrlich, S. D., and Jannière, L. (1993). A fourth class of theta-replicating plasmids: the pAM $\delta$ 1 family from gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11668–11672. doi: 10.1073/pnas.90.24.11668
- Bukowski, M., Lyzen, R., Helbin, W. M., Bonar, E., Szalewska-Palasz, A., Węgrzyn, G., et al. (2013). A regulatory role for the *Staphylococcus aureus* toxin-antitoxin system PemIKSa. *Nat. Commun.* 4, 2012. doi: 10.1038/ncomms3012
- Carr, S. B., Phillips, S. E., and Thomas, C. D. (2016). Structures of replication initiation proteins from staphylococcal antibiotic resistance plasmids reveal



- protein asymmetry and flexibility are necessary for replication. *Nucleic Acids Res.* 44, 2417–2428. doi: 10.1093/nar/gkvl539
- Caryl, J. A., and O'Neill, A. J. (2009). Complete nucleotide sequence of pGO1, the prototype conjugative plasmid from the staphylococci. *Plasmid* 62, 35–38. doi: 10.1016/j.plasmid.2009.03.001
- Catchpole, I. R., and Dyke, K. G. (1992). Replication of staphylococcal plasmid pT48. *FEMS Microbiol. Lett.* 70, 55–62. doi: 10.1111/j.1574-6968.1992.tb05183.x
- Ceglowski, P., Lurz, R., and Alonso, J. C. (1993). Functional analysis of pSM19035-derived replicons in *Bacillus subtilis*. *FEMS Microbiol. Lett.* 109, 145–150. doi: 10.1111/j.1574-6968.1993.tb06159.x
- Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., and Ton-Hoang, B. (2013). Breaking and joining single stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol.* 11, 525–538. doi: 10.1038/nrmicro3067
- Chang, T. L., Naqvi, A., Anand, S. P., Kramer, M. G., Munshi, R., and Khan, S. A. (2002). Biochemical characterization of the *Staphylococcus aureus* PcrA helicase and its role in plasmid rolling circle replication. *J. Biol. Chem.* 277, 45880–45886. doi: 10.1074/jbc.M207383200
- Coffey, A., Harrington, A., Kearney, K., Daly, C., and Fitzgerald, G. (1994). Nucleotide sequence and structural organization of the small, broad-host-range plasmid pCI411 from *Leuconostoc lactis* 533. *Microbiology* 140, 2263–2269. doi: 10.1099/13500872-140-9-2263
- Das, N., and Chatteraj, D. K. (2004). Origin pairing ('hand-cuffing') and unpairing in the control of P1 plasmid replication. *Mol. Microbiol.* 54, 836–849. doi: 10.1111/j.1365-2958.2004.04322.x
- del Solar, G., Diaz, R., and Espinosa, M. (1987). Replication of the streptococcal plasmid pMV158 and derivatives in cell-free extracts of *Escherichia coli*. *Mol. Gen. Genet.* 206, 428–435. doi: 10.1007/BF00428882
- del Solar, G., and Espinosa, M. (1992). The copy number of plasmid pLS1 is regulated by two *trans*-acting plasmid products: the antisense RNAII and the repressor protein, RepA. *Mol. Microbiol.* 6, 83–94. doi: 10.1111/j.1365-2958.1992.tb00840.x
- del Solar, G., and Espinosa, M. (2000). Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* 37, 492–500. doi: 10.1046/j.1365-2958.2000.02005.x
- del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., and Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 62, 434–464.
- Dempsey, L. A., Birch, P., and Khan, S. A. (1992). Six amino acids determine the sequence-specific DNA binding and replication specificity of the initiator proteins of the pT181 family. *J. Biol. Chem.* 267, 24538–24543.
- Dempsey, L. A., Zhao, A. C., and Khan, S. A. (1995). Localization of the start sites of lagging-strand replication of rolling-circle plasmids from gram-positive bacteria. *Mol. Microbiol.* 15, 679–687. doi: 10.1111/j.1365-2958.1995.tb02377.x
- Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapuis, J., Errington, J., et al. (2001). Two essential DNA polymerases at the bacterial replication fork. *Science* 294, 1716–1719. doi: 10.1126/science.1066351
- Diaz, A., Lacks, S. A., and López, P. (1994). Multiple roles for DNA polymerase I in establishment and replication of the promiscuous plasmid pLS1. *Mol. Microbiol.* 14, 773–783. doi: 10.1111/j.1365-2958.1994.tb01314.x
- Diep, B. A., Gill, S. R., Chang, R. F., Phan, T. H., Chen, J. H., Davidson, M. G., et al. (2006). Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet* 367, 731–739. doi: 10.1016/S0140-6736(06)68231-7
- Firth, N., Apisiridej, S., Berg, T., O'Rourke, B. A., Curnock, S., Dyke, K. G., et al. (2000). Replication of staphylococcal multiresistance plasmids. *J. Bacteriol.* 182, 2170–2178. doi: 10.1128/JB.182.8.2170-2178.2000
- Firth, N., and Skurray, R. A. (1998). Mobile elements in the evolution and spread of multiple-drug resistance in staphylococci. *Drug Resist. Updat.* 1, 49–58. doi: 10.1016/S1368-7646(98)80214-8
- Firth, N., and Skurray, R. A. (2006). "Genetics: accessory elements and genetic exchange," in *Gram-positive Pathogens*, 2nd Edn, eds V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (Washington, DC: ASM Press), 413–426.
- Gennaro, M. L., Kornblum, J., and Novick, R. P. (1987). A site-specific recombination function in *Staphylococcus aureus* plasmids. *J. Bacteriol.* 169, 2601–2610. doi: 10.1128/jb.169.6.2601-2610.1987
- Giraldo, R., and Fernández-Tresguerres, M. E. (2004). Twenty years of the pPS10 replicon: insights on the molecular mechanism for the activation of DNA replication in iteron-containing bacterial plasmids. *Plasmid* 52, 69–83. doi: 10.1016/j.plasmid.2004.06.002
- Gomis-Rüth, F. X., Solá, M., Acebo, P., Párraga, A., Guasch, A., Eritja, R., et al. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.* 17, 7404–7415. doi: 10.1093/emboj/17.24.7404
- Goursot, R., Goze, A., Niaudet, B., and Ehrlich, S. D. (1982). Plasmids from *Staphylococcus aureus* replicate in yeast *Saccharomyces cerevisiae*. *Nature* 298, 488–490. doi: 10.1038/298488a0
- Grohmann, E., Goessweiner-Mohr, N., and Brantl, S. (2016). DNA-binding proteins regulating pIP501 transfer and replication. *Front. Mol. Biosci.* 3:42. doi: 10.3389/fmolb.2016.00042
- Gros, M. F., te Riele, H., and Ehrlich, S. D. (1989). Replication origin of a single-stranded DNA plasmid pC194. *EMBO J.* 8, 2711–2716.
- Gruss, A., and Ehrlich, S. D. (1989). The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. *Microbiol. Rev.* 53, 231–241.
- Gruss, A. D., Ross, H. F., and Novick, R. P. (1987). Functional analysis of a palindromic sequence required for normal replication of several staphylococcal plasmids. *Proc. Natl. Acad. Sci. U.S.A.* 84, 2165–2169. doi: 10.1073/pnas.84.8.2165
- Hall, M. C., and Matson, S. W. (1999). Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* 34, 867–877. doi: 10.1046/j.1365-2958.1999.01659.x
- Hernández-Arriaga, A. M., Rubio-Lepe, T. S., Espinosa, M., and del Solar, G. (2009). Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes. *Nucleic Acids Res.* 37, 4799–4811. doi: 10.1093/nar/gkp503
- Ingmer, H., Fong, E. L., and Cohen, S. N. (1995). Monomer-dimer equilibrium of the pSC101 RepA protein. *J. Mol. Biol.* 250, 309–314. doi: 10.1006/jmbi.1995.0378
- Inoue, R., Kaito, C., Tanabe, M., Kamura, K., Akimitsu, N., and Sekimizu, K. (2001). Genetic identification of two distinct DNA polymerases, DnaE and PolC, that are essential for chromosomal DNA replication in *Staphylococcus aureus*. *Mol. Genet. Genomics* 266, 564–571. doi: 10.1007/s004380100564
- Iordanescu, S. (1993). Plasmid pT181-linked suppressors of the *Staphylococcus aureus* pcrA3 chromosomal mutation. *J. Bacteriol.* 175, 3916–3917. doi: 10.1128/jb.175.12.3916-3917.1993
- Iordanescu, S., and Basheer, R. (1991). The *Staphylococcus aureus* mutation pcrA3 leads to the accumulation of pT181 replication initiation complexes. *J. Mol. Biol.* 221, 1183–1189. doi: 10.1016/0022-2836(91)90927-X
- Iordanescu, S., and Surdeanu, M. (1980). New incompatibility groups of *Staphylococcus aureus* plasmids. *Plasmid* 4, 256–260. doi: 10.1016/0147-619X(80)90064-5
- Jannié, L., and Ehrlich, S. D. (1987). Recombination between short repeated sequences is more frequent in plasmids than in the chromosome of *Bacillus subtilis*. *Mol. Gen. Genet.* 210, 116–121. doi: 10.1007/BF00337766
- Khan, S. A. (1997). Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 61, 442–455.
- Khan, S. A., and Novick, R. P. (1982). Structural analysis of plasmid pSN2 in *Staphylococcus aureus*: no involvement in enterotoxin B production. *J. Bacteriol.* 149, 642–649.
- Koepsel, R. R., and Khan, S. A. (1986). Static and initiator protein-enhanced bending of DNA at a replication origin. *Science* 233, 1316–1318. doi: 10.1126/science.3749879
- Konieczny, I., Bury, K., Wawrzycka, A., and Wegrzyn, K. (2014). Iteron Plasmids. *Microbiol. Spectr.* 2:PLAS-0026-2014. doi: 10.1128/microbiolspec.PLAS-0026-2014
- Kornberg, A., and Baker, T. A. (1992). *DNA Replication*, 2nd Edn. New York, NY: W.H. Freeman.
- Kramer, M. G., del Solar, G., and Espinosa, M. (1995). Lagging-strand origins of the promiscuous plasmid pMV158: physical and functional characterization. *Microbiology* 141, 655–662. doi: 10.1099/13500872-141-3-655
- Kramer, M. G., Espinosa, M., Misra, T. K., and Khan, S. A. (1998). Lagging strand replication of rolling-circle plasmids: specific recognition of the ssoA-type origins in different gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10505–10510. doi: 10.1073/pnas.95.18.10505



- Kramer, M. G., Espinosa, M., Misra, T. K., and Khan, S. A. (1999). Characterization of a single-strand origin, *ssoU*, required for broad host range replication of rolling-circle plasmids. *Mol. Microbiol.* 33, 466–475. doi: 10.1046/j.1365-2958.1999.01471.x
- Kramer, M. G., Khan, S. A., and Espinosa, M. (1997). Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. *EMBO J.* 16, 5784–5795. doi: 10.1093/emboj/16.18.5784
- Kwong, S. M., and Firth, N. (2015). Structural and sequence requirements for the antisense RNA regulating replication of staphylococcal multiresistance plasmid pSK41. *Plasmid* 78, 17–25. doi: 10.1016/j.plasmid.2015.01.002
- Kwong, S. M., Lim, R., Lebard, R. J., Skurray, R. A., and Firth, N. (2008). Analysis of the pSK1 replicon, a prototype from the staphylococcal multiresistance plasmid family. *Microbiology* 154, 3084–3094. doi: 10.1099/mic.0.2008/017418-0
- Kwong, S. M., Skurray, R. A., and Firth, N. (2004). *Staphylococcus aureus* multiresistance plasmid pSK41: analysis of the replication region, initiator protein binding and antisense RNA regulation. *Mol. Microbiol.* 51, 497–509. doi: 10.1046/j.1365-2958.2003.03843.x
- Kwong, S. M., Skurray, R. A., and Firth, N. (2006). Replication control of staphylococcal multiresistance plasmid pSK41: an antisense RNA mediates dual-level regulation of Rep expression. *J. Bacteriol.* 188, 4404–4412. doi: 10.1128/JB.00030-06
- Lane, H. E., and Denhardt, D. T. (1975). The *rep* mutation. IV. Slower movement of replication forks in *Escherichia coli rep* strains. *J. Mol. Biol.* 97, 99–112.
- Le Chatelier, E., Jannière, L., Ehrlich, S. D., and Canceille, D. (2001). The RepE initiator is a double-stranded and single-stranded DNA binding protein that forms an atypical open complex at the onset of replication of plasmid pAMB1 from Gram-positive bacteria. *J. Biol. Chem.* 276, 10234–10246. doi: 10.1074/jbc.M010118200
- Lee, C. A., Thomas, J., and Grossman, A. D. (2012). The *Bacillus subtilis* conjugative transposon ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. *J. Bacteriol.* 194, 3165–3172. doi: 10.1128/JB.00301-12
- Leelaporn, A., Firth, N., Byrne, M. E., Roper, E., and Skurray, R. A. (1994). Possible role of insertion sequence IS257 in dissemination and expression of high- and low-level trimethoprim resistance in staphylococci. *Antimicrob. Agents Chemother.* 38, 2238–2244. doi: 10.1128/AAC.38.10.2238
- Leelaporn, A., Firth, N., Paulsen, I. T., Hettiaratchi, A., and Skurray, R. A. (1995). Multidrug resistance plasmid pSK108 from coagulase-negative staphylococci; relationships to *Staphylococcus aureus qacC* plasmids. *Plasmid* 34, 62–67. doi: 10.1006/plas.1995.1034
- Leelaporn, A., Firth, N., Paulsen, I. T., and Skurray, R. A. (1996). IS257-mediated cointegration in the evolution of a family of staphylococcal trimethoprim resistance plasmids. *J. Bacteriol.* 178, 6070–6073. doi: 10.1128/jb.178.20.6070-6073.1996
- Lioy, V. S., Pratto, F., de la Hoz, A. B., Ayora, S., and Alonso, J. C. (2010). Plasmid pSM19035, a model to study stable maintenance in Firmicutes. *Plasmid* 64, 1–17. doi: 10.1016/j.plasmid.2010.04.002
- Liu, M. A., Kwong, S. M., Jensen, S. O., Brzoska, A. J., and Firth, N. (2013). Biology of the staphylococcal conjugative multiresistance plasmid pSK41. *Plasmid* 70, 42–51. doi: 10.1016/j.plasmid.2013.02.001
- Liu, M. A., Kwong, S. M., Pon, C. K., Skurray, R. A., and Firth, N. (2012). Genetic requirements for replication initiation of the staphylococcal multiresistance plasmid pSK41. *Microbiology* 158, 1456–1467. doi: 10.1099/mic.0.057620-0
- López-Aguilar, C., Romero-López, C., Espinosa, M., Berzal-Herranz, A., and del Solar, G. (2015). The 5'-tail of antisense RNAII of pMV158 plays a critical role in binding to the target mRNA and in translation inhibition of *repB*. *Front. Genet.* 6:255. doi: 10.3389/fgene.2015.00225
- López-Aguilar, C., Ruiz-Masó, J. A., Rubio-Lepe, T. S., Sanz, M., and del Solar, G. (2013). Translation initiation of the replication initiator *repB* gene of promiscuous plasmid pMV158 is led by an extended non-SD sequence. *Plasmid* 70, 69–77. doi: 10.1016/j.plasmid.2013.01.011
- Lorenzo-Díaz, F., and Espinosa, M. (2009). Lagging-strand DNA replication origins are required for conjugal transfer of the promiscuous plasmid pMV158. *J. Bacteriol.* 191, 720–727. doi: 10.1128/JB.01257-08
- Lorenzo-Díaz, F., Fernández-López, C., Lurz, R., Bravo, A., and Espinosa, M. (2017). Crosstalk between vertical and horizontal gene transfer: plasmid replication control by a conjugative relaxase. *Nucleic Acids Res.* 45, 7774–7785. doi: 10.1093/nar/gkx450
- Lyon, B. R., and Skurray, R. A. (1987). Antimicrobial resistance of *Staphylococcus aureus*: genetic basis. *Microbiol. Rev.* 51, 88–134.
- Maciag, I. E., Viret, J. F., and Alonso, J. C. (1988). Replication and incompatibility properties of plasmid pUB110 in *Bacillus subtilis*. *Mol. Gen. Genet.* 212, 232–240. doi: 10.1007/BF00334690
- Majumder, S., and Novick, R. P. (1988). Intermediates in plasmid pT181 DNA replication. *Nucleic Acids Res.* 16, 2897–2912. doi: 10.1093/nar/16.7.2897
- Marston, F. Y., Grainger, W. H., Smits, W. K., Hopcroft, N. H., Green, M., Hounslow, A. M., et al. (2010). When simple sequence comparison fails: the cryptic case of the shared domains of the bacterial replication initiation proteins DnaB and DnaD. *Nucleic Acids Res.* 38, 6930–6942. doi: 10.1093/nar/gkq465
- Matsunaga, F., Ishiai, M., Kobayashi, G., Uga, H., Yura, T., and Wada, C. (1997). The central region of RepE initiator protein of mini-F plasmid plays a crucial role in dimerization required for negative replication control. *J. Mol. Biol.* 274, 27–38. doi: 10.1006/jmbi.1997.1373
- Meijer, W. J., van der Lelie, D., Venema, G., and Bron, S. (1995). Effects of the generation of single-stranded DNA on the maintenance of plasmid pMV158 and derivatives in *Lactococcus lactis*. *Plasmid* 33, 91–99. doi: 10.1006/plas.1995.1011
- Mendes, R. E., Deshpande, L. M., Bonilla, H. F., Schwarz, S., Huband, M. D., Jones, R. N., et al. (2013). Dissemination of a pSCFS3-like *cfr*-carrying plasmid in *Staphylococcus aureus* and *Staphylococcus epidermidis* clinical isolates recovered from hospitals in Ohio. *Antimicrob. Agents Chemother.* 57, 2923–2928. doi: 10.1128/AAC.00071-13
- Moscoco, M., del Solar, G., and Espinosa, M. (1995). Specific nicking-closing activity of the initiator of replication protein RepB of plasmid pMV158 on supercoiled or single-stranded DNA. *J. Biol. Chem.* 270, 3772–3779. doi: 10.1074/jbc.270.8.3772
- Müller, A. K., Rojo, F., and Alonso, J. C. (1995). The level of the pUB110 replication initiator protein is autoregulated, which provides an additional control for plasmid copy number. *Nucleic Acids Res.* 23, 1894–1900. doi: 10.1093/nar/23.11.1894
- Needham, C., Noble, W. C., and Dyke, K. G. (1995). The staphylococcal insertion sequence IS257 is active. *Plasmid* 34, 198–205. doi: 10.1006/plas.1995.0005
- Niaudet, B., Jannière, L., and Ehrlich, S. D. (1984). Recombination between repeated DNA sequences occurs more often in plasmids than in the chromosome of *Bacillus subtilis*. *Mol. Gen. Genet.* 197, 46–54. doi: 10.1007/BF00327921
- Nimmo, G. R. (2012). USA300 abroad: global spread of a virulent strain of community-associated methicillin-resistant *Staphylococcus aureus*. *Clin. Microbiol. Infect.* 18, 725–734. doi: 10.1111/j.1469-0691.2012.03822.x
- Noirot, P., Bargonetti, J., and Novick, R. P. (1990). Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8560–8564. doi: 10.1073/pnas.87.21.8560
- Noirot-Gros, M. F., Bidnenko, V., and Ehrlich, S. D. (1994). Active site of the replication protein of the rolling circle plasmid pC194. *EMBO J.* 13, 4412–4420.
- Novick, R. P. (1989). Staphylococcal plasmids and their replication. *Annu. Rev. Microbiol.* 43, 537–563. doi: 10.1146/annurev.mi.43.100189.002541
- Novick, R. P., Iordanescu, S., Projan, S. J., Kornblum, J., and Edelman, I. (1989). pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell* 59, 395–404. doi: 10.1016/0092-8674(89)90300-0
- Novick, R. P., Projan, S. J., Rosenblum, W., and Edelman, I. (1984). Staphylococcal plasmid cointegrates are formed by host- and phage-mediated general *rec* systems that act on short regions of homology. *Mol. Gen. Genet.* 195, 374–377. doi: 10.1007/BF00332777
- O'Brien, F. G., Yui Eto, K., Murphy, R. J., Fairhurst, H. M., Coombs, G. W., Grubb, W. B., et al. (2015). Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in

- Staphylococcus aureus*. *Nucleic Acids Res.* 43, 7971–7983. doi: 10.1093/nar/gkv755
- Oliveira, S. S., Murphy, E., Gamon, M. R., and Bastos, M. C. (1993). pRJ5: a naturally occurring *Staphylococcus aureus* plasmid expressing constitutive macrolide-lincosamide-streptogramin B resistance contains a tandem duplication in the leader region of the *ermC* gene. *J. Gen. Microbiol.* 139, 1461–1467. doi: 10.1099/00221287-139-7-1461
- Pérez-Roth, E., Kwong, S. M., Alcoba-Florez, J., Firth, N., and Méndez-Alvarez, S. (2010). Complete nucleotide sequence and comparative analysis of pPR9, a 41.7-kilobase conjugative staphylococcal multiresistance plasmid conferring high-level mupirocin resistance. *Antimicrob. Agents Chemother.* 54, 2252–2257. doi: 10.1128/AAC.01074-09
- Petit, M. A., Dervyn, E., Rose, M., Entian, K. D., McGovern, S., Ehrlich, S. D., et al. (1998). PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol. Microbiol.* 29, 261–273. doi: 10.1046/j.1365-2958.1998.00927.x
- Planet, P. J. (2017). Life after USA300: The rise and fall of a superbug. *J. Infect. Dis.* 215(Suppl. 1), S71–S77. doi: 10.1093/infdis/jiw444
- Polard, P., Marsin, S., McGovern, S., Veltou, M., Wigley, D. B., Ehrlich, S. D., et al. (2002). Restart of DNA replication in Gram-positive bacteria: functional characterisation of the *Bacillus subtilis* PriA initiator. *Nucleic Acids Res.* 30, 1593–1605. doi: 10.1093/nar/30.7.1593
- Priebe, S. D., and Lacks, S. A. (1989). Region of the streptococcal plasmid pMV158 required for conjugative mobilisation. *J. Bacteriol.* 171, 4778–4784. doi: 10.1128/jb.171.9.4778-4784.1989
- Puyet, A., del Solar, G. H., and Espinosa, M. (1988). Identification of the origin and direction of replication of the broad-host-range plasmid pLS1. *Nucleic Acids Res.* 16, 115–133. doi: 10.1093/nar/16.1.115
- Ramsay, J. P., and Firth, N. (2017). Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.* 38, 1–9. doi: 10.1016/j.mib.2017.03.003
- Ramsay, J. P., Kwong, S. M., Murphy, R. J., Yui Eto, K., Price, K. J., Nguyen, Q. T., et al. (2016). An updated view of plasmid conjugation and mobilization in *Staphylococcus*. *Mob. Genet. Elements* 6:e1208317. doi: 10.1080/2159256X.2016.1208317
- Rasooly, A., and Novick, R. P. (1993). Replication-specific inactivation of the pT181 plasmid initiator protein. *Science* 262, 1048–1050. doi: 10.1126/science.8235621
- Rasooly, A., Wang, P. Z., and Novick, R. P. (1994). Replication-specific conversion of the *Staphylococcus aureus* pT181 initiator protein from an active homodimer to an inactive heterodimer. *EMBO J.* 13, 5245–5251.
- Ruby, C., and Novick, R. P. (1975). Plasmid interactions in *Staphylococcus aureus*: nonadditivity of compatible plasmid DNA pools. *Proc. Natl. Acad. Sci. U.S.A.* 72, 5031–5035. doi: 10.1073/pnas.72.12.5031
- Ruiz-Masó, J. A., Anand, S. P., Espinosa, M., Khan, S. A., and del Solar, G. (2006). Genetic and biochemical characterization of the *Streptococcus pneumoniae* PcrA helicase and its role in plasmid rolling circle replication. *J. Bacteriol.* 188, 7416–7425. doi: 10.1128/JB.01010-06
- Ruiz-Masó, J. A., Bordanaba-Ruiseño, L., Sanz, M., Menéndez, M., and del Solar, G. (2016). Metal-induced stabilization and activation of plasmid replication initiator RepB. *Front. Mol. Biosci.* 3:56. doi: 10.3389/fmolb.2016.00056
- Ruiz-Masó, J. A., López-Zumel, C., Menéndez, M., Espinosa, M., and del Solar, G. (2004). Structural features of the initiator of replication protein RepB encoded by the promiscuous plasmid pMV158. *Biochim. Biophys. Acta* 1696, 113–119. doi: 10.1016/j.bbapap.2003.09.010
- Ruiz-Masó, J. A., Lurz, R., Espinosa, M., and del Solar, G. (2007). Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication. *Nucleic Acids Res.* 35, 1230–1244. doi: 10.1093/nar/gkl1099
- Ruiz-Masó, J. A., Machón, C., Bordanaba-Ruiseño, L., Espinosa, M., Coll, M., and del Solar, G. (2015). Plasmid rolling-circle replication. *Microbiol. Spectr.* 3:PLAS-0035-2014. doi: 10.1128/microbiolspec.PLAS-0035-2014
- Sanders, G. M., Dallmann, H. G., and McHenry, C. S. (2010). Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol. Cell* 37, 273–281. doi: 10.1016/j.molcel.2009.12.025
- Schumacher, M. A., Tonthat, N. K., Kwong, S. M., Chinnam, N. B., Liu, M. A., Skurray, R. A., et al. (2014). Mechanism of staphylococcal multiresistance plasmid replication origin assembly by the RepA protein. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9121–9126. doi: 10.1073/pnas.1406065111
- Schwarz, S., Shen, J., Wendlandt, S., Fessler, A. T., Wang, Y., Kadlec, K., et al. (2014). Plasmid-mediated antimicrobial resistance in staphylococci and other firmicutes. *Microbiol. Spectr.* 2:PLAS-0020-2014. doi: 10.1128/microbiolspec.PLAS-0020-2014
- Seegers, J. F., Zhao, A. C., Meijer, W. J., Khan, S. A., Venema, G., and Bron, S. (1995). Structural and functional analysis of the single-strand origin of replication from the lactococcal plasmid pWV01. *Mol. Gen. Genet.* 249, 43–50. doi: 10.1007/BF00290234
- Simpson, A. E., Skurray, R. A., and Firth, N. (2000). An IS257-derived hybrid promoter directs transcription of a *tetA(K)* tetracycline resistance gene in the *Staphylococcus aureus* chromosomal *mec* region. *J. Bacteriol.* 182, 3345–3352. doi: 10.1128/JB.182.12.3345-3352.2000
- Slatter, A. F., Thomas, C. D., and Webb, M. R. (2009). PcrA helicase tightly couples ATP hydrolysis to unwinding double-stranded DNA, modulated by the initiator protein for plasmid replication, RepD. *Biochemistry* 48, 6326–6334. doi: 10.1021/bi900101h
- Smits, W. K., Merrikh, H., Bonilla, C. Y., and Grossman, A. D. (2011). Primosomal proteins DnaD and DnaB are recruited to chromosomal regions bound by DnaA in *Bacillus subtilis*. *J. Bacteriol.* 193, 640–648. doi: 10.1128/JB.01253-10
- Soultanas, P., Dillingham, M. S., Papadopoulos, F., Phillips, S. E., Thomas, C. D., and Wigley, D. B. (1999). Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase. *Nucleic Acids Res.* 27, 1421–1428. doi: 10.1093/nar/27.6.1421
- Steinmetzer, K., Behlke, J., and Brantl, S. (1998). Plasmid pIP501 encoded transcriptional repressor CopR binds to its target DNA as a dimer. *J. Mol. Biol.* 283, 595–603. doi: 10.1006/jmbi.1998.2122
- Stewart, P. R., Dubin, D. T., Chikramane, S. G., Inglis, B., Matthews, P. R., and Poston, S. M. (1994). IS257 and small plasmid insertions in the *mec* region of the chromosome of *Staphylococcus aureus*. *Plasmid* 31, 12–20. doi: 10.1006/plas.1994.1002
- Swan, M. K., Bastia, D., and Davies, C. (2006). Crystal structure of pi initiator protein-iteron complex of plasmid R6K: implications for initiation of plasmid DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18481–18486. doi: 10.1073/pnas.0609046103
- Takahashi, S., Hours, C., Chu, A., and Denhardt, D. T. (1979). The *rep* mutation. VI. Purification and properties of the *Escherichia coli* rep protein, DNA helicase III. *Can. J. Biochem.* 57, 855–866. doi: 10.1139/o79-105
- Thomas, C. D., Balson, D. F., and Shaw, W. V. (1990). *In vitro* studies of the initiation of staphylococcal plasmid replication. Specificity of RepD for its origin (*oriD*) and characterization of the Rep-*ori* tyrosyl ester intermediate. *J. Biol. Chem.* 265, 5519–5530.
- Thomas, C. D., Nikiforov, T. T., Connolly, B. A., and Shaw, W. V. (1995). Determination of sequence specificity between a plasmid replication initiator protein and the origin of replication. *J. Mol. Biol.* 254, 381–391. doi: 10.1006/jmbi.1995.0625
- Toukdarian, A. E., and Helinski, D. R. (1998). TrfA dimers play a role in copy-number control of RK2 replication. *Gene* 223, 205–211. doi: 10.1016/S0378-1119(98)00370-9
- Udo, E. E., and Grubb, W. B. (1991). A new incompatibility group plasmid in *Staphylococcus aureus*. *FEMS Microbiol. Lett.* 62, 33–36. doi: 10.1111/j.1574-6968.1991.tb04412.x
- Udo, E. E., Wei, M. Q., and Grubb, W. B. (1992). Conjugative trimethoprim resistance in *Staphylococcus aureus*. *FEMS Microbiol. Lett.* 76, 243–248. doi: 10.1111/j.1574-6968.1992.tb05470.x
- Wang, P. Z., Projan, S. J., Henriquez, V., and Novick, R. P. (1992). Specificity of origin recognition by replication initiator protein in plasmids of the pT181 family is determined by a six amino acid residue element. *J. Mol. Biol.* 223, 145–158. doi: 10.1016/0022-2836(92)90722-V
- Wassenaar, T. M., Ussery, D. W., and Ingmer, H. (2016). The *qacC* gene has recently spread between rolling circle plasmids of *Staphylococcus*, indicative of a novel gene transfer mechanism. *Front. Microbiol.* 7:1528. doi: 10.3389/fmicb.2016.01528
- Weaver, K. E., Kwong, S. M., Firth, N., and Francia, M. V. (2009). The RepA\_N replicons of Gram-positive bacteria: a family of broadly distributed but

- narrow host range plasmids. *Plasmid* 61, 94–109. doi: 10.1016/j.plasmid.2008.11.004
- Weisblum, B., Graham, M. Y., Gryczan, T., and Dubnau, D. (1979). Plasmid copy number control: isolation and characterization of high-copy-number mutants of plasmid pE194. *J. Bacteriol.* 137, 635–643.
- Yamaguchi, T., Hayashi, T., Takami, H., Ohnishi, M., Murata, T., Nakayama, K., et al. (2001). Complete nucleotide sequence of a *Staphylococcus aureus* exfoliative toxin B plasmid and identification of a novel ADP-ribosyltransferase, EDIN-C. *Infect. Immun.* 69, 7760–7771. doi: 10.1128/IAI.69.12.7760-7771.2001
- Zhang, W., Dillingham, M. S., Thomas, C. D., Allen, S., Roberts, C. J., and Soutanas, P. (2007). Directional loading and stimulation of PcrA helicase by the replication initiator protein RepD. *J. Mol. Biol.* 371, 336–348. doi: 10.1016/j.jmb.2007.05.050
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2017 Kwong, Ramsay, Jensen and Firth. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Biofilm-Forming Clinical *Staphylococcus* Isolates Harbor Horizontal Transfer and Antibiotic Resistance Genes

Sandra Águila-Arcos<sup>1</sup>, Itxaso Álvarez-Rodríguez<sup>1</sup>, Olatz Garaiurrebaso<sup>1</sup>, Carlos Garbisu<sup>2</sup>, Elisabeth Grohmann<sup>3</sup> and Itziar Alkorta<sup>1\*</sup>

<sup>1</sup> Instituto Biofisika (UPV/EHU, CSIC), Department of Biochemistry and Molecular Biology, University of the Basque Country, Bilbao, Spain, <sup>2</sup> Department of Conservation of Natural Resources, Soil Microbial Ecology Group, NEIKER-Tecnalia, Derio, Spain, <sup>3</sup> Life Sciences and Technology, Beuth University of Applied Sciences, Berlin, Germany

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Guenther Muth,  
Universität Tübingen, Germany  
Fabián Lorenzo,  
Universidad de La Laguna, Spain  
Gloria Del Solar,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### \*Correspondence:

Itziar Alkorta  
itzi.alkorta@ehu.eus

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 21 June 2017

**Accepted:** 02 October 2017

**Published:** 16 October 2017

### Citation:

Águila-Arcos S, Álvarez-Rodríguez I,  
Garaiurrebaso O, Garbisu C,  
Grohmann E and Alkorta I (2017)  
Biofilm-Forming Clinical  
*Staphylococcus* Isolates Harbor  
Horizontal Transfer and Antibiotic  
Resistance Genes.  
Front. Microbiol. 8:2018.  
doi: 10.3389/fmicb.2017.02018

Infections caused by staphylococci represent a medical concern, especially when related to biofilms located in implanted medical devices, such as prostheses and catheters. Unfortunately, their frequent resistance to high doses of antibiotics makes the treatment of these infections a difficult task. Moreover, biofilms represent a hot spot for horizontal gene transfer (HGT) by bacterial conjugation. In this work, 25 biofilm-forming clinical staphylococcal isolates were studied. We found that *Staphylococcus epidermidis* isolates showed a higher biofilm-forming capacity than *Staphylococcus aureus* isolates. Additionally, horizontal transfer and relaxase genes of two common staphylococcal plasmids, pSK41 and pT181, were detected in all isolates. In terms of antibiotic resistance genes, *aac6-aph2a*, *ermC*, and *tetK* genes, which confer resistance to gentamicin, erythromycin, and tetracycline, respectively, were the most prevalent. The horizontal transfer and antibiotic resistance genes harbored on these staphylococcal clinical strains isolated from biofilms located in implanted medical devices points to the potential risk of the development and dissemination of multiresistant bacteria.

**Keywords:** Staphylococci, biofilm, relaxases, antibiotic resistance, nosocomial infections

## INTRODUCTION

Staphylococci, mainly *Staphylococcus aureus* and *Staphylococcus epidermidis*, are well-known causative agents of a large number of human infectious diseases, including skin, soft tissue, respiratory tract, bone, joint and endovascular infections, as well as infections related to implanted medical devices (Otto, 2012; Le et al., 2014). Their pathogenicity is due not only to the virulence factors that they express, but also to the ability of these bacteria to form biofilms (i.e., deeply seated microbial communities attached to inert or living surfaces; Costerton et al., 1999; Otto, 2008). The treatment of biofilm-associated infections is considered a challenging task owing to their inherent resistance to (i) antimicrobial agents and (ii) the host immune system (Hoiby et al., 2010). Moreover, nowadays, the incidence of antibiotic resistant pathogenic bacteria in clinical settings is



dramatically increasing, making treatment of bacterial infections one of our most serious health threats (Guridi et al., 2015). This problem arises from the resistance phenotype of bacteria that harbor resistance genes in their chromosomal and/or plasmid DNA.

Bacteria can acquire resistance genes by horizontal gene transfer (HGT). Actually, conjugative plasmid-mediated HGT is considered the most important process in the emergence of new resistant pathogens (Schiwon et al., 2013). It is well-documented that bacterial conjugation can occur within biofilms since they provide an ideal situation for the exchange of genetic material of various origins (Christensen et al., 1998; Hausner and Wuertz, 1999). On the other hand, bacterial conjugation can induce biofilm formation since the cell-to-cell contact established for gene exchange favors the close proximity of bacteria required for biofilm formation (Ghigo, 2001; Molin and Tolker-Nielsen, 2003; Reisner et al., 2006; Yang et al., 2008; D'Alvise et al., 2010). This link between biofilms and bacterial conjugation increases both the risk of biofilm-related infections and the conjugative spread of virulence factors.

In this work, we studied 25 staphylococcal biofilm-forming clinical isolates belonging to the following species: *S. aureus*, *S. epidermidis*, *S. hominis*, and *S. capitis*. These species are commonly found on human skin and can cause biofilm-forming healthcare-associated infections. Both horizontal transfer and antibiotic resistance genes were detected in these staphylococcal clinical isolates. This work adds valuable information on the risk of development and dissemination of antibiotic resistance in *Staphylococcus* biofilm-forming clinical isolates.

## MATERIALS AND METHODS

### Bacterial Strains

A total of 25 staphylococcal biofilm-forming clinical isolates were kindly provided by Hospital Universitario Donostia, Spain. In addition, they provided data on their antibiotic resistance phenotype, determined by diffusion discs on agar. The origin and antibiotic resistance phenotype of each isolate are shown in **Table 1**.

### Growth Conditions

Swabs from the clinical isolates were plated on tryptic soy agar (TSA) and incubated at 37°C. Subsequently, a single colony of each isolate was grown in 10 ml of tryptic soy broth (TSB) supplemented with at least two antibiotics to which the strain was phenotypically resistant (see **Table 1**), at 37°C overnight. The culture was centrifuged at 8,000 × g for 10 min. Then, the pellet was resuspended in 2 ml of TSB medium containing 40% (v/v) glycerol and stored at −80°C.

For this study, strains were grown in TSB medium at 37°C and 200 rpm. TSB medium and TSA plates were supplemented, when required, with amoxicillin (8 µg/ml), cloxacillin [2 µg/ml for *S. aureus* and 0.5 µg/ml for coagulase negative staphylococci (CoNS)], erythromycin (4 µg/ml), mupirocin (520 µg/ml), tetracycline (8 µg/ml), gentamicin (20 µg/ml), rifampicin (2 µg/ml), or levofloxacin (2 µg/ml).

### DNA Extraction

Plasmid DNA was extracted from the 25 clinical isolates with the ATP™ Plasmid Midi kit (ATP biotech Inc., Taiwan), according to the manufacturer's instructions.

### Detection of Small Plasmids by Agarose Gel Electrophoresis

To detect small plasmids (molecular size < 20 kb), 1 µg of total extracted plasmid DNA was linearized by incubation with 30 U of *Aspergillus oryzae* nuclease S1 (Sigma, Spain) at 37°C for 45 min. Nuclease S1 cuts one strand of the DNA at the nick site and its activity results in linearized plasmids (Germond et al., 1974). Different enzyme concentrations were studied to optimize nuclease S1 digestion (data not shown). Linearized plasmids were visualized on 1% (w/v) agarose gels in 1 × TAE buffer.

### Detection of Large Plasmids by Pulsed Field Gel Electrophoresis

Detection of large plasmids (molecular size > 20 kb) was carried out by Pulsed Field Gel Electrophoresis (PFGE) as described by Barton et al. (1995) with modifications. Bacteria were grown in 2 ml of TSB medium overnight at 37°C and 200 rpm. Cultures were diluted in PIV buffer [10 mM Tris-HCl (pH 8), 1 M NaCl] until OD<sub>600</sub> = 1. Then, 600 µl of diluted culture were centrifuged at 11,000 × g for 2 min. Subsequently, the pellet was washed with 500 µl of PIV buffer and centrifuged again. The pellet was resuspended in 300 µl of PIV buffer and incubated at 42°C for 10 min. Next, 150 µl of the sample were mixed with 150 µl of 2% (w/v) low-melting agarose (BioRad) which had been preincubated at 42°C. The mixture was transferred into the plugs, incubated at room temperature for 10 min and, subsequently, for 15 min at 4°C. Once solidified, gel plugs were incubated at 37°C for 5–6 h with shaking (600 rpm) in 1 ml of lysis buffer EC [6 mM Tris-HCl (pH 8), 1 M NaCl, 100 mM EDTA (pH 8), 0.2% (w/v) sodium deoxycholate, 0.5% (w/v) n-lauroylsarcosine, 100 µg/ml lysozyme, 50 µg/ml lysostaphin]. After cell lysis, gel plugs were transferred to new tubes containing 1 ml of EPS solution [1% (w/v) n-lauroylsarcosine, 0.5 M EDTA (pH 8), 100 µg/ml proteinase K] and then incubated at 56°C for 16–20 h. Next, five washes with 1 ml of TE buffer [10 mM Tris-HCl (pH 8), 1 mM EDTA (pH 8)] at 50°C for 30 min each were carried out. For nuclease S1 digestion, each gel plug was cut into two slices. Each slice was incubated twice in 100 µl of digestion solution [50 mM NaCl, 30 mM sodium acetate (pH 4.5), 5 mM ZnSO<sub>4</sub>] at room temperature for 15 min. Then, slices were incubated at 37°C for 45 min with 1 U of *A. oryzae* nuclease S1 (Sigma) in 100 µl of digestion solution. The reaction was stopped by transferring the slices to 1 ml of TE buffer for 1 h. Digested slices were applied to wells in 1% (w/v) Pulsed Field Certified Agarose (BioRad) prepared in 0.5 × TBE buffer [45 mM Tris (pH 8), 45 mM boric acid, 1 mM EDTA] and run in CHEF-DR® III System (BioRad) at 6 V/cm, a field angle of 120°, and switch times of 5 to 35 s for 22 h. Lambda Ladder PFGE (New England Biolabs, Ipswich, U.S.) was used as molecular size marker and pSK41 plasmid (46.4 kb) was used as positive control. Gels were stained with GelRed Nucleic Acid Stain (Biogen Científica,

**TABLE 1** | Origin and antibiotic resistance phenotype of the Staphylococcal biofilm-forming clinical isolates used in this work.

No.	Isolate	Origin	Antibiotic resistance <sup>a,b</sup>
1	<i>S. aureus</i> 312042	Prosthesis	AMX, AMC, CFZ, CLI, CLOX, ERY, LVX
2	<i>S. aureus</i> 410099	Prosthesis	AMX, AMC, CFZ, CLOX, GEN, LVX, MUP
3	<i>S. aureus</i> 218154	Prosthesis	AMX, AMC, CFZ, CLI, CLOX, ERY, LVX, MUP, RIF
4	<i>S. aureus</i> 339031	Catheter	AMX, AMC, CFZ, CLI, CLOX, ERY, LVZ
5	<i>S. aureus</i> 215642	Prosthesis	AMX
6	<i>S. epidermidis</i> 213303	Prosthesis	AMX, AMC, CFZ, CLI, CLOX, CTX, ERY, GEN, LVX, MUP, RIF
7	<i>S. hominis</i> 313732	Prosthesis	AMX, AMC, CFZ, CLI, CLOX, CTX, ERY, GEN, LVX, MUP
8	<i>S. capitis</i> 316479	Prosthesis	AMX, AMC, CFZ, CLOX, GEN, LVX, MUP
9	<i>S. epidermidis</i> 319622	Prosthesis	AMX, ERY, TET
10	<i>S. epidermidis</i> 219691	Prosthesis	AMX, ERY, LVX, MUP, RIF
11	<i>S. aureus</i> 214967	Ulcer	AMX, MUP
12	<i>S. epidermidis</i> 239879	Catheter	AMX, AMC, CFZ, CLI, CLOX, ERY, GEN, MUP, TET
13	<i>S. epidermidis</i> 239891	Catheter	AMX, AMC, CFZ, CLOX, GEN, LVX, MUP, RIF
14	<i>S. aureus</i> 337423-1	Catheter	AMX, AMC, CFZ, CLOX, LVX
15	<i>S. aureus</i> 338550-1	Catheter	AMX, AMC, CFZ, CLOX, ERY, LVX
16	<i>S. aureus</i> 339031-2	Catheter	AMX, AMC, CFZ, CLOX, ERY, LVX
17	<i>S. aureus</i> 339056-2	Catheter	AMX, AMC, CFZ, CLOX, ERY, LVX
18	<i>S. aureus</i> 339300	Catheter	AMX, AMC, CFZ, CLOX, LVX
19	<i>S. aureus</i> 338503	Catheter	AMX
20	<i>S. epidermidis</i> 214627-A	Articular fluid from patient with prosthesis	AMX, AMC, CFZ, CLI, CLOX, CTX, ERY, GEN, LVX, MUP, RIF
21	<i>S. epidermidis</i> 310301-1	Articular fluid from patient with prosthesis	AMX, AMC, CFZ, CLI, CLOX, CTX, ERY, GEN, LVX, MUP, RIF
22	<i>S. epidermidis</i> 338400-1	Catheter	AMX, AMC, CFZ, CLOX, MUP
23	<i>S. epidermidis</i> 338515-1	Catheter	AMX, AMC, CFZ, CLI, CLOX, CTX, ERY, GEN, MUP, RIF
24	<i>S. epidermidis</i> 338684	Catheter	AMX, AMC, CFZ, CLI, CLOX, CTX, GEN, LVX, MUP, RIF
25	<i>S. epidermidis</i> 216663	Articular fluid from patient with prosthesis	AMX, AMC, CFZ, CLI, CLOX, ERY, GEN, RIF

<sup>a</sup>Resistance to antibiotics was analyzed by diffusion discs on agar by Hospital Universitario Donostia.

<sup>b</sup>AMX, amoxicillin; AMC, amoxicillin + clavulanic acid; CFZ, cefazolin; CLI, clindamycin; CLOX, cloxacillin; CTX, cotrimoxazol; ERY, erythromycin; GEN, gentamicin; LVX, levofloxacin; MUP, mupirocin; RIF, rifampicin; TET, tetracycline; VAN, vancomycin.

Madrid, Spain). Bands were visualized by ChemiDoc XRS System (BioRad). Images were analyzed by Quantity One 1-D Analysis Software (BioRad).

## Polymerase Chain Reaction (PCR) and Southern Blotting

PCR and Southern blotting assays, specific for horizontal transfer and antibiotic resistance genes, were performed using the strains and plasmids indicated in **Table 2** as reference DNA. Oligonucleotides used for gene detection are listed in **Table 3**. Each 25 µl PCR reaction mixture contained 1.25 U Taq polymerase (New England Biolabs, Ipswich, U.S.), 1 × PCR buffer, 0.5 µM of each primer, 0.2 mM deoxynucleoside triphosphates and 20 ng of template DNA (plasmid DNA). Amplifications were carried out in a C1000<sup>TM</sup> Thermal Cycler (BioRad). PCR temperature profiles are shown in **Table 4**. PCR products were separated by agarose gel electrophoresis, transferred to a membrane (Sambrook and Russel, 2001), and then hybridized with the corresponding specific DIG-labeled probe using the PCR DIG Probe Synthesis Kit (Roche, Mannheim, Germany). Detection of DNA sequences was performed with the DIG Luminescent Detection Kit (Roche) according to the manufacturer's instructions.

**TABLE 2** | Bacterial strains and plasmids used as reference for PCR and Southern blotting.

Strain	Characteristics	References
<b><i>Bacillus subtilis</i></b>		
BD662	pBD90, <i>ermD</i>	Gryczan et al., 1984
BD1156	pBD370, <i>ermG</i>	Monod et al., 1987
<b><i>Enterococcus faecalis</i></b>		
RE25	pRE25, <i>ermB</i> , <i>tetM</i>	Schwarz et al., 2001
V583	pTEF1, pTEF2, pTEF3, <i>vanB</i>	Paulsen et al., 2003
<b><i>S. aureus</i></b>		
RN3259	pT181, <i>tetK</i> , <i>pre</i> <sub>pT181</sub>	Khan et al., 1981
SK5428	pSK41, <i>acc(6')-Ie-aph(2')-Ia</i> , <i>tra</i> <sup>+</sup> , <i>pre</i> <sub>pSK41</sub> , <i>neS</i> <sub>pSK41</sub>	Firth et al., 1993
<b><i>S. haemolyticus</i></b>		
VPS617	<i>tetK</i> , <i>ermC</i>	Perreten et al., 2005

## Biofilm Formation

To test the 25 clinical isolates for biofilm formation, a quantitative adherence assay (Christensen et al., 1985) with some modifications was used. Briefly, 200 µl of TSB medium

**TABLE 3 |** Oligonucleotides used for the detection of antibiotic resistance and transfer genes.

Gene	Oligonucleotide	Sequence (5' → 3')	Acc. No <sup>a</sup>	Nucleotide position	Reference
<b>Antibiotic resistance genes</b>	<i>aac6-aph2a</i> fw	GCCAGAACATGAATTACACGAG	NC_005024	42,981–43,002	Schiwon et al., 2013
	<i>aac6-aph2a</i> rev	CTGTTGTTGCATTAGTCTTTCC		43,569–43,591	
	<i>ermB</i> fw	GCATTTAACGACGAACTGGCT	U00453	6,796–6,817	Böckelmann et al., 2009
	<i>ermB</i> rev	GACAATACTTGCTCATAAGTAATGGT		7,343–7,368	
	<i>ermC</i> fw	CGTAAGTCCATTGAAATAGACC	V01278	2,555–2,577	Schiwon et al., 2013
	<i>ermC</i> rev	TCCTGCATGTTTTAAGGAATTG		2,079–2,100	
	<i>ermD</i> fw	CGGGCAATATTAGCATAGACG	M29832	544–565	Schiwon et al., 2013
	<i>ermD</i> rev	ATTCTGACCATTGCCGAGTC		988–1,007	
	<i>ermG</i> fw	TGCAGGGAAAGGTCATTTTAC	M15332	785–805	Schiwon et al., 2013
	<i>ermG</i> rev	AACCCATTTTATTACAAAAGTTTC		1,245–1,268	
	<i>tetK</i> fw	TTTGAGCTGTCTTGTTTCATTG	CP000045	1,398–1,419	Schiwon et al., 2013
	<i>tetK</i> rev	AGCCCACCAGAAAACAAACC		1,918–1,937	
	<i>tetM</i> fw	GAAGTCGAACAAGAGGAAAGC	M85225	1,114–1,134	Tenover and Rasheed, 2004
	<i>tetM</i> rev	ATGGAAGCCCAGAAAAGGAT		1,835–1,853	
	<i>vanB</i> fw	CCCGAATTTCAAATGATTGAAAA	L06138	113–135	Miele et al., 1995
	<i>vanB</i> rev	CGCCATCCTCCTGCAAAA		552–569	
<b>Transfer genes</b>	<i>pSK41 pre</i> fw	CTGGACTAAAGGCATGCAA	AF051917	20,674–20,693	Schiwon et al., 2013
	<i>pSK41 pre</i> rev	GCAGTTTTCCATCACGCATA		20,298–20,317	
	<i>pSK41 nes</i> fw	AGCGCTAGTAGGATTAAG	AF051917	10,016–10,034	Schiwon et al., 2013
	<i>pSK41 nes</i> rev	CATAATAATGTGCGTGAGG		9,706–9,725	
	<i>pT181 pre</i> fw	TCGAACAGAATTATACAGGCAA	CP000045	2,708–2,729	Schiwon et al., 2013
	<i>pT181 pre</i> rev	CTGACTTATTTGCTCATGTTTACG		3,082–3,105	
	<i>traE</i> pSK41 fw	TATCATGATCC(T/C)GAA(A/G)ATGAAT	AF051917	27,456–27,478	Schiwon et al., 2013
	<i>traE</i> pSK41 rev	TCTTTTGT(T/G)ATTTCTGCCATAA		28,060–28,082	
	<i>traG</i> pSK41 fw	GTGTTGACGGTTCGGGTATC	AF051917	30,132–30,151	Schiwon et al., 2013
	<i>traG</i> pSK41 rev	TTTTCCGTCTGAACCTCCAC		30,570–30,589	
	<i>traK</i> pSK41 fw	TATCTAAAGACCACCCAGCTAGAG	AF051917	34,636–34,660	Schiwon et al., 2013
	<i>traK</i> pSK41 rev	TACTTGTTTCAAACCTCTACAGTAGC		35,185–35,209	
	<i>traL</i> pSK41 fw	ATGGGGACTATGGCAGGTAG	AF051917	36,279–36,298	Schiwon et al., 2013
	<i>traL</i> pSK41 rev	AAGTTTTTGACCACTTCCAG		36,667–36,686	
	<i>traM</i> pSK41 fw	TGTTGTATGGGAAAACAAGC	AF051917	36,870–36,890	Schiwon et al., 2013
	<i>traM</i> pSK41 rev	GCTGGGCTTATAGC(A/G)TCATC		37,051–37,070	

<sup>a</sup>Accession Number from Gene Bank.**TABLE 4 |** PCR conditions.

Genes amplified	Denaturation	Primer annealing	Elongation
<i>aac6-aph2a</i> , <i>ermD</i> , <i>ermG</i> ,	95°C, 30 s	55°C, 30 s	72°C, 30 s
<i>ermC</i> , <i>tetK</i> , <i>tetM</i>	95°C, 30 s	58°C, 30 s	72°C, 30 s
<i>ermB</i> , <i>vanB</i>	95°C, 30 s	60°C, 45 s	72°C, 60 s
<i>traE</i> , <i>traG</i> , <i>traK</i> , <i>traL</i> , <i>traM</i> , <i>pre<sub>pT181</sub></i>	95°C, 30 s	55°C, 45 s	72°C, 60 s
<i>pre<sub>pSK41</sub></i> , <i>nes<sub>pSK41</sub></i>	95°C, 60 s	50°C, 60 s	72°C, 120 s

An initial denaturation step was performed, consisting of 2 min at 95°C, except for *pre<sub>pSK41</sub>*, *nes<sub>pSK41</sub>*, that were denaturated for 4 min. Then, 30 cycles consisting of denaturation, primer annealing and elongation steps were performed at the conditions (temperature and time) specified. A final elongation step at 72°C was performed during 5 min, except for *pre<sub>pSK41</sub>*, *nes<sub>pSK41</sub>* in which it lasted 10 min.

in 96-well flat-bottom polystyrene plates were inoculated with 10 µl overnight bacterial cultures and grown without shaking at 37°C for 24 h. Planktonic bacteria were removed from

each well. Then, three washes with distilled water per well were carried out. Next, 125 µl of 0.1% (w/v) crystal violet solution were added to each well and incubated for 10 min at room temperature. Subsequently, three washes with distilled water were again performed. To solubilize the dye, 200 µl of 33% (v/v) glacial acetic acid solution were added to each stained well and incubated for 10 min at room temperature. TSB medium was used as negative control. The optical density of the attached bacteria was measured in a microplate reader at 570 nm (in triplicate for each strain). The ability to form biofilm was attributed as: OD<sub>570</sub> < 0.120, no biofilm-forming; 0.120 < OD<sub>570</sub> < 0.240, weak biofilm-forming; OD<sub>570</sub> > 0.240, strong biofilm-forming (Christensen et al., 1985; Di Rosa et al., 2006); and OD<sub>570</sub> > 1.5, very strong biofilm-forming. Dilutions were performed when absorbance values were higher than the limit of accurate detection. To classify the isolates into significant groups, statistical analysis was performed using SigmaPlot program and Student's *t*-test or Mann–Whitney *U*-test.

## RESULTS

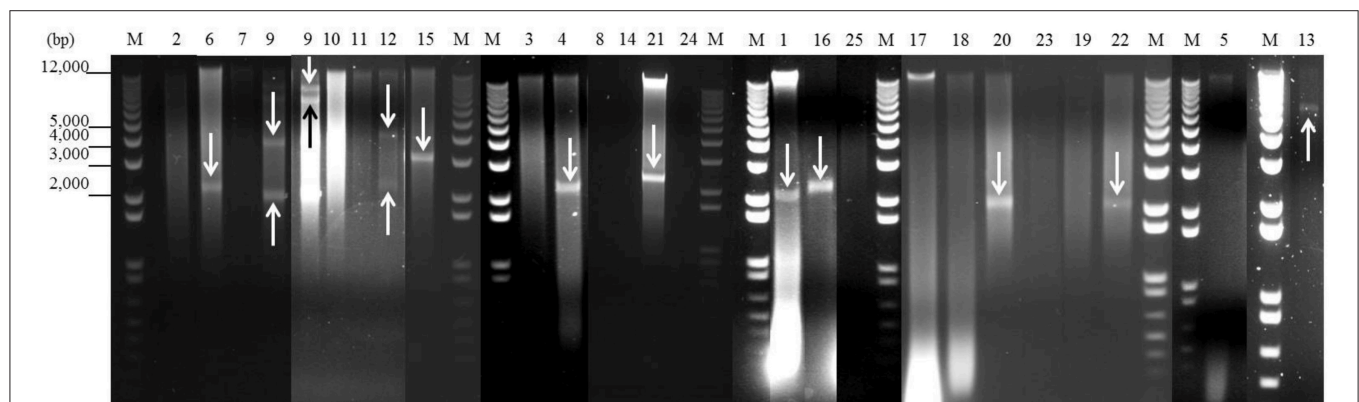
### All Clinical Isolates, Except One, Harbored Plasmids

Plasmid DNA was extracted from the 25 clinical isolates and then analyzed by agarose gel electrophoresis and PFGE (Figures 1, 2). Since plasmid DNA samples are sometimes contaminated with chromosomal DNA, as suggested in Figure 1 for some of our isolates (i.e., 1, 3, 17, 21), after the extraction of plasmid DNA, we decided to test for such contamination. To this purpose, 16S rRNA from isolates 1, 3, 17, and 21 was amplified by PCR as explained in Broszat et al. (2014). The obtained amplicons were analyzed by 1% (w/v) agarose gel electrophoresis in 1 × TAE buffer. As observed in Supplementary Figure 1, some of our plasmid DNA samples appear to be contaminated with chromosomal DNA. Nonetheless, as reflected

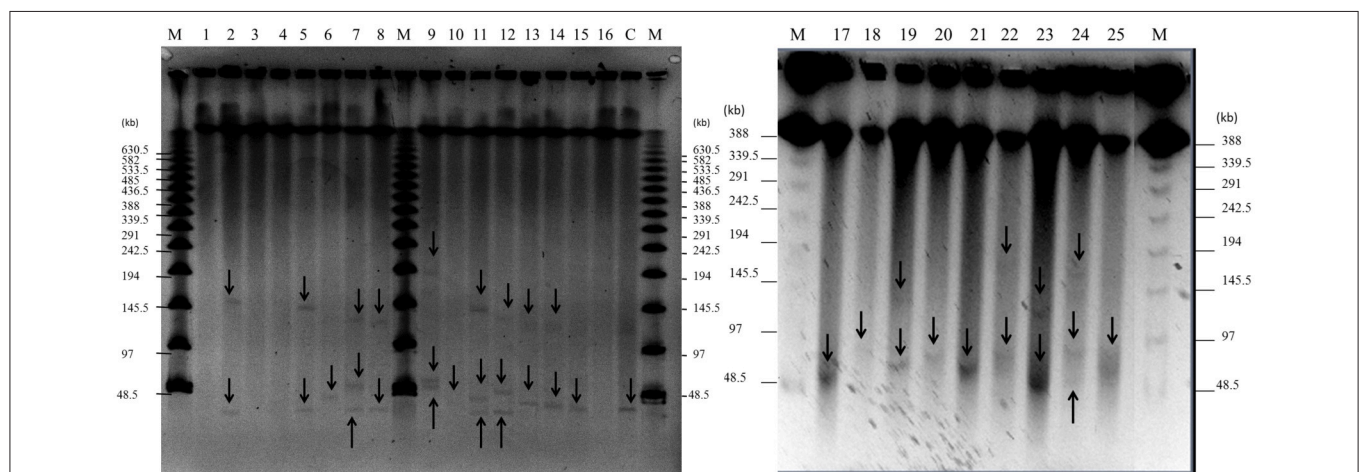
in Figures 1, 2, the majority of the extracted DNA corresponds to plasmid DNA.

As shown in Figures 1, 2 and Table 5, a total of 54 plasmids of sizes ranging from 2 to 200 kb were detected using both methods: 15 small plasmids (size < 20 kb) and 39 large plasmids (size > 20 kb; Shearer et al., 2011). All clinical isolates contained at least one plasmid, except isolate 3. The combination of agarose gel electrophoresis and PFGE is unable to detect plasmids between 13 and 45 kb. Then, *a priori*, our clinical isolates could harbor more plasmids than observed here. In particular, isolate 3 could harbor a plasmid between 13 and 45 kb, which could explain the apparent lack of plasmid observed for this isolate.

When agarose gel electrophoresis was used, it was observed that 44% of the clinical isolates contained at least one plasmid with a size < 20 kb (Figure 1 and Table 5). In particular, nine of the isolates contained only one plasmid smaller than 20 kb.



**FIGURE 1 |** Detection of plasmids from 25 staphylococcal clinical isolates by agarose gel electrophoresis after digestion with nuclease S1. One microgram of plasmid DNA from each isolate was digested with 30 U of nuclease S1 at 37°C for 45 min. After digestion, the plasmids were analyzed by 1% (w/v) agarose gel electrophoresis in 1 × TAE buffer. Lanes 1–25: digested plasmid DNA from each strain (lane numbers correspond to the number of the isolate). Lanes M: DNA molecular weight marker 1 kb Plus DNA Ladder. Bands corresponding to plasmids are indicated with arrows.



**FIGURE 2 |** Detection of plasmids in 25 staphylococcal clinical isolates by PFGE. Large plasmids (>30 kb) were analyzed by PFGE after digestion with nuclease S1 at 37°C for 45 min. Lanes 1–25: clinical isolates; lane number corresponds to the number of the isolate. Lane C: positive control, plasmid pSK41 (46.4 kb) extracted from SK5428 strain. Lane M: Lambda Ladder PFGE molecular size marker. Arrows point to detected plasmids.



**TABLE 5 |** Antibiotic resistance profiles, transfer genes, plasmid content, and biofilm-forming capacity of staphylococcal clinical isolates.

Isolate	Antibiotic resistance phenotype <sup>a</sup>	Antibiotic resistance genotype <sup>b</sup>	Transfer genes	Plasmids <sup>c</sup>		Biofilm-forming capacity 24 h <sup>d</sup>
				<20 kb	>20 kb	
1	ERY	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traE, traG, traK, traL, traM*</i>	1	0	0
2	GEN	<i>ermB, ermC, ermG, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traG, traL</i>	0	2	1
3	ERY	<i>ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traL*</i>	0	0	1
4	ERY	<i>ermB*, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traL</i>	1	0	0
5	–	<i>ermB, ermC, tetK*, aac6-aph2a*</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traK, traL</i>	0	2	3
6	ERY, GEN	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traG, traK, traL</i>	1	1	2
7	ERY, GEN	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traE, traK, traL</i>	0	3	2
8	GEN	<i>ermC, tetK*, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traG, traK</i>	0	2	3
9	ERY, TET	<i>ermC, tetK*</i>	<i>pre<sub>PSK41</sub>*, pre<sub>PT181</sub>, traL*</i>	4	3	2
10	ERY	<i>ermB, ermC, tetK, aac6-aph2a, vanB</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traG, traL</i>	0	1	2
11	–	<i>ermB, ermC, tetK</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traE, traG, traL, traM</i>	0	3	1
12	ERY, GEN, TET	<i>ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>*, pre<sub>PT181</sub>, traG*, traL</i>	2	3	3
13	GEN	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traE*, traG*, traK, traL*, traM*</i>	1	2	2
14	–	<i>ermC, tetK*, aac6-aph2a</i>	<i>pre<sub>PT181</sub>, traE, traG*, traK, traL</i>	0	2	0
15	ERY	<i>ermB, ermC, ermG*, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traL*</i>	1	1	1
16	ERY	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traE*, traG*, traK, traL</i>	1	0	0
17	ERY	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traE, traG, traK, traL, traM</i>	0	1	2
18	–	<i>ermB*, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, nes<sub>PSK41</sub>, pre<sub>PT181</sub>, traE, traG, traK, traL</i>	0	1	0
19	–	<i>ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traK*, traM</i>	0	2	2
20	ERY, GEN	<i>ermB*, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traE*, traG*, traK*, traL, traM</i>	1	1	2
21	ERY, GEN	<i>ermB*, ermC, tetK</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traE, traG, traK, traL, traM</i>	1	1	2
22	–	<i>ermB, ermC, tetK, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traE*, traG*, traL, traM</i>	1	2	3
23	ERY, GEN	<i>ermC, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traE*, traG, traL</i>	0	2	2
24	GEN	<i>ermB*, ermC, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub></i>	0	3	2
25	ERY, GEN	<i>ermB, ermC, tetK, tetM, aac6-aph2a</i>	<i>pre<sub>PSK41</sub>, pre<sub>PT181</sub>, traG*, traK*, traL, traM</i>	0	1	3

<sup>a</sup> GEN, gentamicin; ERY, erythromycin; TET, tetracycline.

<sup>b</sup> *aac6-aph2a*, gentamicin; *ermB/ermC/ermG*, erythromycin; *tetK/tetM*, tetracycline; *vanB*, vancomycin resistance genes.

<sup>c</sup> Numbers indicate the number of plasmid bands observed in the 1% agarose gel or in the PFGE.

<sup>d</sup> 0, no biofilm-forming capacity; 1, weak biofilm-forming capacity; 2, strong biofilm-forming capacity; 3, very strong biofilm-forming capacity.

\*Weak signal intensity in the Southern blot.

Isolate 12 harbored two plasmids smaller than 20 kb, while 4 plasmids of this size were identified in isolate 9.

According to our PFGE data, 84% of the clinical isolates (all except isolates 1, 3, 4, and 16) contained at least one large plasmid (Figure 2 and Table 5): 32% of the isolates (6, 10, 15, 17, 18, 20, 21, and 25) harbored one large plasmid; 32% of the isolates (2, 5, 8, 13, 14, 19, 22, and 23) contained two large plasmids; and 20% of the isolates (7, 9, 11, 12, and 24) harbored three large plasmids.

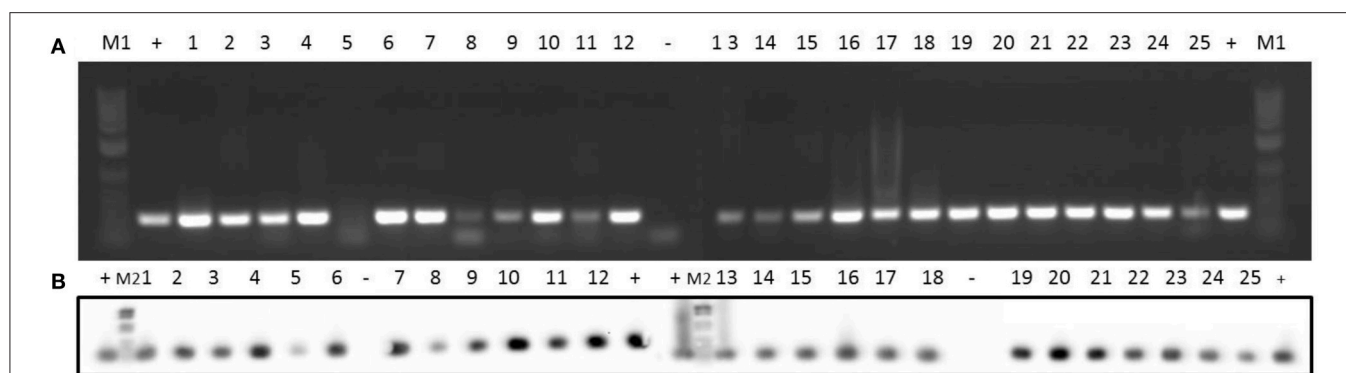
## All Clinical Isolates Contained Antibiotic Resistance Genes

Eight resistance genes commonly found in staphylococci were investigated by PCR and Southern blotting: genes encoding resistance to erythromycin (*ermB*, *ermC*, *ermD*, *ermG*), tetracycline (*tetK*, *tetM*), gentamicin (*aac6-aph2a*), and vancomycin (*vanB*). The presence of these genes was tested in our extracted DNA (i.e., putative plasmid DNA) because, initially, we were only interested in the risk of dissemination of antibiotic resistance from these clinical strains through bacterial conjugation.

Concerning erythromycin resistance, 15 of the strains had an erythromycin resistance phenotype (Table 1). Data at the genotype level for the different clinical isolates are shown in Table 5. *ermC* gene was observed in all the isolates (Figure 3), while *ermD* was not detected in any of the isolates (Supplementary Figure 2). Likewise, 72% of the isolates were *ermB*-positive (Supplementary Figure 3), whereas only 8% of the isolates (2 and 15) harbored the *ermG* gene (Supplementary Figure 4).

With respect to tetracycline, only two isolates (9 and 12) were observed to be tetracycline resistant at the phenotype level (Table 1). Regarding this antibiotic, 23 out of 25 isolates contained the *tetK* gene (Supplementary Figure 5), while only isolate 25 harbored the *tetM* gene (Supplementary Figure 6).

Similarly, the 25 isolates were analyzed for the occurrence of the gentamicin resistance *aac6-aph2a* gene. As shown in Supplementary Figure 7, this gene was detected in 88% of the isolates (all the isolates except 9, 11, and 21 showed a positive result for the *aac6-aph2a* gene). However, according to the phenotype (Table 1), only 44% of the isolates showed gentamicin resistance.



**FIGURE 3 |** Detection of erythromycin resistance gene *ermC* in 25 clinical isolates by PCR (A) and Southern blotting (B). Amplicons of *ermC* (477 bp) were visualized on 1% (w/v) agarose gels. Lanes 1–25: clinical isolates. Lanes +: positive control. Lanes -: negative control. Lanes M1: DNA molecular weight marker 1 kb Plus DNA Ladder. Lanes M2: DNA molecular weight marker VI DIG-labeled.

Finally, regarding vancomycin resistance, all the isolates were phenotypically sensitive to this antibiotic (Table 1). At the genotype level, only isolate 10 proved to be *vanB*-positive (Figure 4).

### All Clinical Isolates Encoded Relaxase and/or Horizontal Transfer Genes Commonly Found in *Staphylococcus* Conjugative/Mobilizable Plasmids

In order to find out whether the abovementioned antibiotic resistance genes were likely to be disseminated via conjugative transfer, we searched for horizontal transfer genes from two common staphylococcal plasmids: (i) conjugative pSK41 and (ii) mobilizable pT181 (Novick, 1989; Berg et al., 1998).

In relation to pSK41, the *pre* relaxase gene was found in all the isolates except isolate 14 (Supplementary Figure 8). In addition, isolates 1, 4, 7, 8, 11, 16, 17, and 18 contained the *nes* relaxase gene of pSK41 (Supplementary Figure 9). Five genes (*traE*, *traG*, *traK*, *traL*, and *traM*) from the transfer region of pSK41 were also analyzed: *traE* gene was present in 48% of the isolates (Supplementary Figure 10), *traG* gene was detected in 68% of the isolates (Supplementary Figure 11), *traK* gene was found in 56% of the isolates (Figure 5), and *traL* gene was detected in 88% of the isolates (Supplementary Figure 12). Finally, *traM* gene was found in only 36% of the isolates (Supplementary Figure 13).

In addition, we tested for the presence of the *pre* relaxase gene of the staphylococcal mobilizable plasmid pT181. As shown in Figure 6, this gene was detected in all the clinical isolates.

### Clinical Isolates Differed in their Biofilm-Forming Capacity

All the clinical strains were isolated from biofilms formed on medical devices such as catheters and prostheses, as well as from ulcer and articular fluids from patients with prostheses (Table 1). In order to confirm their biofilm forming capacity, we used the *in vitro* assay described above (Christensen et al., 1985; Di Rosa et al., 2006).

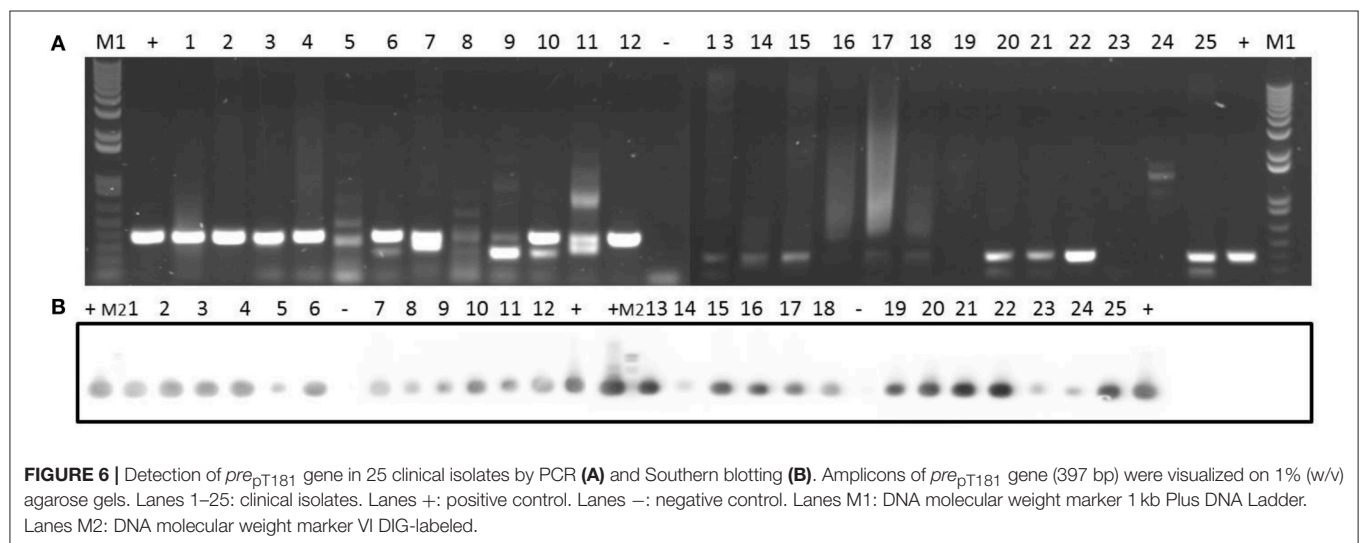
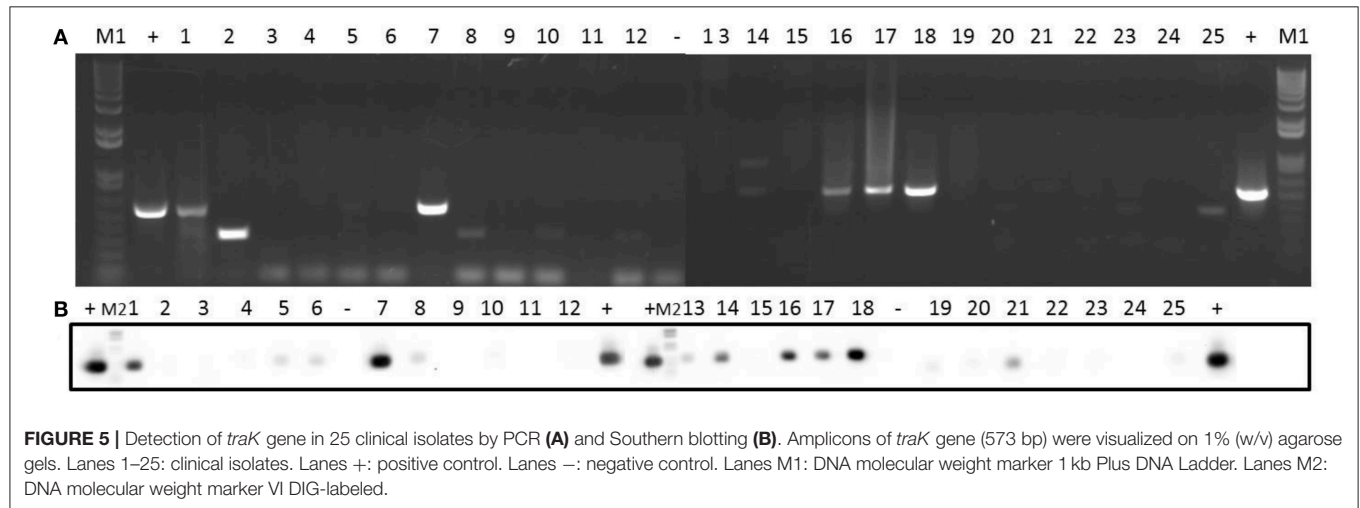
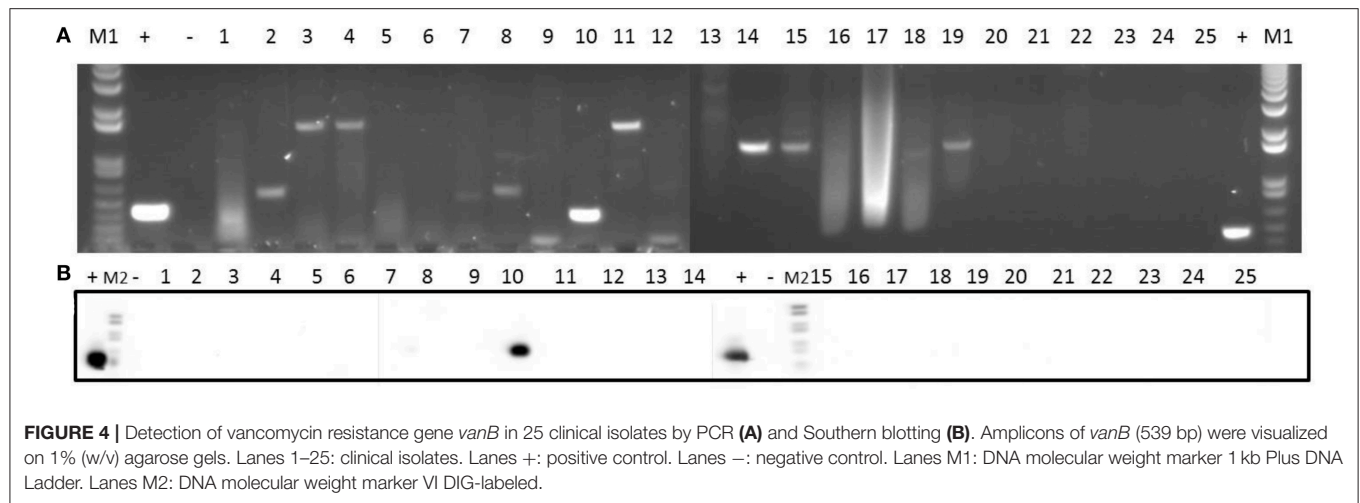
As shown in Figure 7, isolates were divided into four groups: (i) no biofilm-forming isolates: 1, 4, 14, 16, 18; (ii) weak biofilm-forming isolates: 2, 3, 11, 15; (iii) strong biofilm-forming isolates: 6, 7, 9, 10, 13, 17, 19, 20, 21, 23, 24; and (iv) very strong biofilm-forming isolates: 5, 8, 12, 22, 25. Additionally, our statistical analysis showed that the “strong biofilm-forming” group could be further divided into three different sub-groups with increasing biofilm forming capacity from “strong biofilm-forming (1)” to “strong biofilm-forming (3).” The distribution of isolates in these three sub-groups was as follows: “strong biofilm-forming (1)”: 7, 10, 17, 23; “strong biofilm-forming (2)”: 9, 19; and “strong biofilm-forming (3)”: 6, 13, 20, 21, 24 (Figure 7). Furthermore, the relationship between biofilm-forming capacity and *Staphylococcus* species was studied. *S. epidermidis* isolates had a significantly higher ( $p < 0.001$ ) biofilm-forming capacity than *S. aureus* isolates.

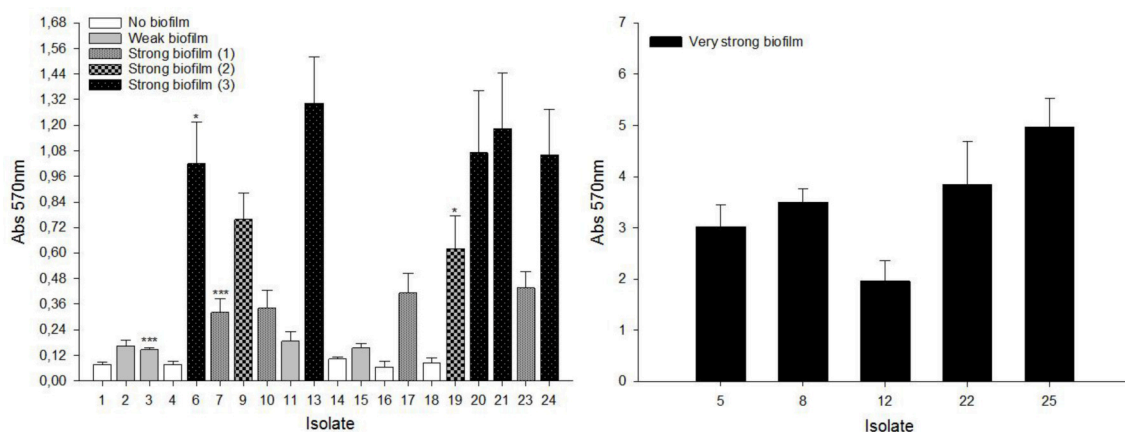
## DISCUSSION

*Staphylococci* nosocomial pathogens are frequently involved in biomaterial-associated infections (Pfaffler and Herwaldt, 1988; Kloos and Bannerman, 1994; Huebner and Goldmann, 1999; Otto, 2008). The eradication of these biofilm-associated infections with antibiotic treatment is usually impossible without the removal of the medical device (Stewart and Costerton, 2001; Mack et al., 2004; Otto, 2012; Tong et al., 2012). Furthermore, conjugative plasmid-mediated dissemination of antibiotic resistance is favored in bacterial biofilms (Ghigo, 2001; Molin and Tolker-Nielsen, 2003; Reisner et al., 2006; Yang et al., 2008; D’Alvise et al., 2010).

In this work, 25 staphylococcal biofilm-forming clinical isolates were studied. First, plasmids of different sizes were detected. Secondly, antibiotic resistance and transfer genes were detected by PCR and Southern blotting. Finally, the capacity of these isolates to form biofilms *in vitro* was studied.

Fifteen plasmids smaller than 20 kb and 39 plasmids larger than 20 kb were found in 11 (44%) and 21 (84%) isolates, respectively. This higher percentage of isolates with large





**FIGURE 7 |** Biofilm-forming capacity of staphylococcal clinical isolates. 96-well flat-bottom polystyrene plates were incubated for 24 h at 37°C without shaking. Cells attached to the wells were stained with 0.1% (w/v) crystal violet. Absorption at 570 nm was measured to quantify biofilm formation. Results are the mean  $\pm$  SEM of at least 4 independent biological experiments performed in triplicate. To classify the isolates into significant groups, statistical analysis was performed using Student's *t*-test or Mann-Whitney *U*-test (\* $p < 0.05$ , \*\*\* $p < 0.001$ ).

plasmids, compared to isolates with small plasmids, is in agreement with results obtained by Shearer et al. (2011) who found that 79% of their isolates harbored at least one large (>20 kb) plasmid. According to Smillie et al. (2010), in proteobacteria, 58% of the plasmids larger than 20 kb are mobilizable. Therefore, our results suggest that almost all our *Staphylococcus* clinical isolates could harbor conjugative and/or mobilizable plasmids.

On the other hand, the presence of antibiotic resistance and horizontal transfer genes commonly found staphylococci was investigated. Antibiotic sensitivity tests (to obtain the well-known antibiograms) are the most common method to determine antibiotic resistance of pathogenic bacteria. Nonetheless, the study of antibiotic resistance at the genotype level is crucial to get information on the potential of those bacteria to develop and disseminate resistance against antibiotics (Palmer and Kishony, 2013). Erythromycin, tetracyclines, gentamicin, and vancomycin are the most used antibiotics for the treatment of staphylococcal infections, but resistance genes against these antibiotics have been described in staphylococcal clinical isolates (Duran et al., 2012; Emameini et al., 2013). Therefore, we searched in our 25 clinical isolates for the presence of genes involved in the resistance to these antibiotics.

Macrolide antibiotics, such as erythromycin, are broad-spectrum antibiotics; relevantly, anti-biofilm activities have been assigned to them (Parra-Ruiz et al., 2012; Zhao et al., 2015). In this work, 60% of the isolates were phenotypically resistant to erythromycin. At the genotype level, different studies have reported the prevalence of the *ermC* gene in staphylococci (Duran et al., 2012; Schiwon et al., 2013); in our study, the *ermC* gene was identified in all the clinical isolates. Although a low prevalence of the *ermB* gene has been described in *Staphylococcus* (Zmantar et al., 2011), in our study, 72% of the isolates harbored the *ermB* gene.

After penicillin, tetracyclines are the second most widely used group of antibiotics worldwide (van Hoek et al., 2011). Resistance

to tetracycline can be encoded in plasmid-located *tet* genes such as *tetK* and *tetL*, or, alternatively, in genes located in the chromosome or transposons such as *tetM* and *tetO* (Emameini et al., 2013). A high incidence of the *tetK* gene (92%) was observed here, whereas only one isolate contained the *tetM* gene. Several studies have reported the coexistence of both *tetM* and *tetK* genes in staphylococci strains (Duran et al., 2012; Camoez et al., 2013; Emameini et al., 2013; Schiwon et al., 2013). Here, a disagreement between phenotypic and genotypic data was observed, since only 8% of the isolates were phenotypically resistant to tetracycline.

Aminoglycosides, such as gentamicin, are broad-spectrum antibiotics used against *S. aureus* infections. Aminoglycoside modifying enzymes (AME) are used by bacteria to abolish the effect of these antibiotics. In *S. aureus* strains, one of the most common genes encoding AME is the *aac6-aph2a* gene (Emameini et al., 2013). In our study, 88% of the isolates contained this gene, in agreement with other studies on staphylococcal isolates (Duran et al., 2012; Emameini et al., 2013). In terms of the phenotype, 44% of the isolates showed resistance to gentamicin (Table 1). Similar discrepancies between phenotypic and genotypic results have been observed by other authors (Duran et al., 2012; Emameini et al., 2013).

The lack of correlation between resistance phenotypic and genotypic data could be due to mutations in genes resulting in non-functional proteins, as well as to the lack of gene expression (Martineau et al., 2000). Also, methods to detect antibiotic resistance phenotype are influenced by technical variables such as temperature, incubation time, inoculum density and so on (Baddour et al., 2007). Likewise, the pattern of negative resistance phenotype together with a positive resistance genotype can be due to the presence of pseudogenes (Davis et al., 2011). As a consequence, it is essential to take this fact into account because it indicates that bacteria have the potential to be resistant to more antibiotics than those shown phenotypically.



Vancomycin has been used to treat staphylococcal infections, mainly methicillin resistant *S. aureus* (Huebner and Goldmann, 1999). In the late 1980s, the emergence of vancomycin resistance was reported for the first time (van Hoek et al., 2011). One of the genes responsible for vancomycin resistance is the *vanB* gene (van Hoek et al., 2011), which was only found in isolate 10. This low incidence of the *vanB* gene, together with the fact that all isolates were phenotypically sensitive to vancomycin (Table 1), suggest that (i) vancomycin is still one of the best options for the treatment of staphylococcal infections and (ii) it should be then used judiciously.

Concerning the presence of horizontal transfer genes, plasmid pSK41 is a prototypical multiresistance plasmid of 46 kb from *S. aureus* (Berg et al., 1998). Therefore, we searched for *pre*<sub>pSK41</sub> and *nes*<sub>pSK41</sub> genes, as well as for five different *tra* genes involved in the conjugative transfer of plasmid pSK41, in our clinical isolates. Although all the isolates, except for one, contained the *pre*<sub>pSK41</sub> gene, only 32% of the isolates harbored the *nes*<sub>pSK41</sub> gene. In addition, 20% of the isolates contained the five *tra* genes tested here. According to these results, and taking into account that isolates 11 and 17 harbored plasmids of around 46 kb, we speculate that these strains could contain pSK41-type plasmids. Other studies have identified plasmids of the pSK41 family in geographically diverse isolates of both *S. aureus* and CoNS (Berg et al., 1998). The coexistence of *pre*<sub>pSK41</sub> and *nes*<sub>pSK41</sub> genes, together with *tra* genes, in some of our isolates, points to a risk of dissemination of resistance traits.

pT181 plasmid is also common among staphylococci (Khan and Novick, 1983; Novick, 1989). Then, we tested for the presence of the relaxase *pre*<sub>pT181</sub> gene responsible for pT181 mobilization. All isolates harbored the *pre*<sub>pT181</sub> relaxase gene, suggesting that pT181-type plasmids could be present in all of the samples. pT181 is a low copy number plasmid and then, not surprisingly, we could not detect it in the electrophoretic gels; however, it may be present at undetectable amounts in some strains. This is of great concern especially in those strains where potentially conjugative plasmids that could mobilize these small plasmids are present.

Finally, in general, *S. epidermidis* isolates have shown a higher biofilm-forming capacity than *S. aureus* isolates. Although, the 25 isolates studied here were obtained from biofilms present in the clinical environment, some of the *S. aureus* isolates were unable to form biofilms under our experimental conditions. This is probably because biofilms in clinical conditions take longer times to form, in comparison to the standardized *in vitro* method used here, in which 24 h was the biofilm-forming time.

The fact that our clinical isolates contained both antibiotic resistance and horizontal transfer genes, as well as conjugative and/or mobilizable plasmids, suggest the possibility of their disseminating antibiotic resistance to other bacteria. Here, it must be stated that, due to the abovementioned presence of chromosomal DNA in our extracted DNA samples, we cannot rule out the possibility that the antibiotic resistance genes identified here were encoded in the chromosomal DNA. However, as reflected in Figures 1, 2, the majority of the extracted DNA corresponds to plasmid DNA. For example, the high incidence of the plasmid-encoded *tetK* gene in our isolates could support this fact. In any case, genes encoded

in the chromosome can also be mobilized between bacterial cells. For instance, transposons can mobilize chromosomal genes by jumping into plasmids or phages which can then be transferred into other cells (Frost et al., 2014). On the other hand, the conjugation process can also occur via chromosomally integrated conjugative elements, such as conjugative transposons. Integrated conjugative elements are known to encode proteins that facilitate their own transfer and sometimes the transfer of other cellular DNA from the donor (Frost et al., 2014). Indeed, as reported by Wilkins and Frost (2001), many plasmids and integrated conjugative elements can effect the transfer of chromosomal DNA. Then, if some of the antibiotic resistance genes identified here were encoded in the chromosomal DNA present in some of our samples, the risk of transfer to other bacterial cells would still exist, although *a priori* lower than if they were encoded in the observed plasmids.

Recent studies underline the importance of collecting more epidemiological data on antibiotic resistance, in order to design novel control strategies for this growing global health problem (Frieri et al., 2016). The isolation and molecular characterization of plasmids from nosocomial pathogens will provide valuable information in the search for new strategies to control the dissemination of antibiotic resistance among clinical pathogens.

## AUTHOR CONTRIBUTIONS

SÁ and IÁR: Acquisition of the data, writing and revision of the content, approval of the last version of the work. OG: Revision of the content, approval of the last version of the work. CG: Writing and revision of the content, approval of the last version and ensuring accuracy and integrity of the work. EG: Design of the work, revision of the content, approval of the last version, and ensuring accuracy and integrity of the work. IA: Design of the work and the acquisition of the data, writing, and revision of the content, approval of the last version and ensuring accuracy and integrity of the work.

## ACKNOWLEDGMENTS

This work was financially supported by the Spanish Ministry of Economy (Grant No BFU2012-36241) and MICINN (Grant No. BFU2010-22103). SÁ and OG were graduate students supported by the Basque Government and Fundación Biofísica Bizkaia. At the moment, IÁR is a graduate student supported by the Basque Government. We thank Eneritz Bilbao for excellent technical assistance. We thank Prof. Pérez-Trallero and Dr. Alonso from the Hospital Universitario Donostia for providing the clinical isolates, and Prof. Quindós and Aketza Varona from the University of the Basque Country, and Dr. Rodríguez-Lázaro and Dr. Hernández from the Instituto Tecnológico Agrario de Castilla y León, Spain for assistance with the PFGE.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02018/full#supplementary-material>

# REFERENCES

- Baddour, M. M., AbuElKheir, M. M., and Fatani, A. J. (2007). Comparison of mecA polymerase chain reaction with phenotypic methods for the detection of methicillin-resistant *Staphylococcus aureus*. *Curr. Microbiol.* 55, 473–479. doi: 10.1007/s00284-007-9015-6
- Barton, B. M., Harding, G. P., and Zuccarelli, A. J. (1995). A general method for detecting and sizing large plasmids. *Anal. Biochem.* 226, 235–240. doi: 10.1006/abio.1995.1220
- Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A., Leelaporn, A., and Skurray, R. A. (1998). Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. *J. Bacteriol.* 180, 4350–4359.
- Böckelmann, U., Dorries, H., Ayuso-Gabella, M. N., de Salgot Marçay, M., Tandoi, V., Levantesi, C., et al. (2009). Quantitative PCR monitoring of antibiotic resistance genes and bacterial pathogens in three European artificial groundwater recharge systems. *Appl. Environ. Microbiol.* 75, 154–163. doi: 10.1128/AEM.01649-08
- Broszat, M., Nacke, H., Blasi, R., Siebe, C., Huebner, J., Daniel, R., et al. (2014). Wastewater irrigation increases the abundance of potentially harmful gammaproteobacteria in soil in Mezquital Valley, Mexico. *Appl. Environ. Microbiol.* 80, 5282–5291. doi: 10.1128/AEM.01295-14
- Camoez, M., Sierra, J. M., Pujol, M., Hornero, A., Martin, R., and Domínguez, M. A. (2013). Prevalence and molecular characterization of methicillin-resistant *Staphylococcus aureus* ST398 resistant to tetracycline at a Spanish Hospital over 12 years. *PLoS ONE* 8:e72828. doi: 10.1371/journal.pone.0072828
- Christensen, B. B., Sternberg, C., Andersen, J. B., Eberl, L., Moller, S., Givskov, M., et al. (1998). Establishment of new genetic traits in a microbial biofilm community. *Appl. Environ. Microbiol.* 64, 2247–2255.
- Christensen, G. D., Simpson, W. A., Younger, J. J., Baddour, L. M., Barrett, F. F., Melton, D. M., et al. (1985). Adherence of coagulase-negative staphylococci to plastic tissue culture plates: a quantitative model for the adherence of staphylococci to medical devices. *J. Clin. Microbiol.* 22, 996–1006
- Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science* 284, 1318–1322. doi: 10.1126/science.284.5418.1318
- D’Alvise, P. W., Sjöholm, O. R., Yankelevich, T., Jin, Y., Wuertz, S., and Smets, B. F. (2010). TOL plasmid carriage enhances biofilm formation and increases extracellular DNA content in *Pseudomonas putida* KT2440. *FEMS Microbiol. Lett.* 312, 84–92. doi: 10.1111/j.1574-6968.2010.02105.x
- Davis, M. A., Besser, T. E., Orfe, L. H., Baker, K. N., Lanier, A. S., Broschat, S. L., et al. (2011). Genotypic-phenotypic discrepancies between antibiotic resistance characteristics of *Escherichia coli* isolates from calves in management settings with high and low antibiotic use. *Appl. Environ. Microbiol.* 77, 3293–3299. doi: 10.1128/AEM.02588-10
- Di Rosa, R., Creti, R., Venditti, W., D’Amelio, R., Arciola, C. R., Montanaro, L., et al. (2006). Relationship between biofilm formation, the enterococcal surface protein and gelatinase in clinical isolates of *Enterococcus faecalis* and *Enterococcus faecium*. *FEMS Microbiol. Lett.* 256, 145–150. doi: 10.1111/j.1574-6968.2006.00112.x
- Duran, N., Ozer, B., Duran, G. G., Onlen, Y., and Demir, C. (2012). Antibiotic resistance genes and susceptibility patterns in staphylococci. *Indian J. Med. Res.* 135, 389–396.
- Emanini, M., Bigverdi, R., Kalantar, D., Soroush, S., Jabalameli, F., Noorazar Khoshnab, B., et al. (2013). Distribution of genes encoding tetracycline resistance and aminoglycoside modifying enzymes in *Staphylococcus aureus* strains isolated from a burn center. *Ann. Burns Fire Disasters* 26, 76–80.
- Firth, N., Ridgway, K. P., Byrne, M. E., Fink, P. D., Johnson, L., Paulsen, I. T., et al. (1993). Analysis of a transfer region from the staphylococcal conjugative plasmid pSK41. *Gene* 136, 13–25. doi: 10.1016/0378-1119(93)90442-6
- Frieri, M., Kumar, K., and Boutin, A. (2016). Antibiotic resistance. *J. Infect. Public Health* 10, 369–378. doi: 10.1016/j.jiph.2016.08.007
- Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2014). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. doi: 10.1038/nrmicro1235
- Germond, J. E., Vogt, V. M., and Hirt, B. (1974). Characterization of the single-strand-specific nuclease S1 activity on double-stranded supercoiled polyoma DNA. *Eur. J. Biochem.* 43, 591–600. doi: 10.1111/j.1432-1033.1974.tb03446.x
- Ghigo, J. M. (2001). Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412, 442–445. doi: 10.1038/35086581
- Gryczan, T., Israeli-Reches, M., Del Bue, M., and Dubnau, D. (1984). DNA sequence and regulation of ermD, a macrolide-lincosamidestreptogramin B resistance element from *Bacillus licheniformis*. *Mol. Gen. Genet.* 194, 349–356. doi: 10.1007/BF00425543
- Guridi, A., Diederich, A. K., Águila-Arcos, S., García-Moreno, M., Blasi, R., Broszat, M., et al. (2015). New antimicrobial contact catalyst killing antibiotic resistant clinical and waterborne pathogens. *Mater. Sci. Eng. C Mater. Biol. Appl.* 50, 1–11. doi: 10.1016/j.msec.2015.01.080
- Hausner, M., and Wuertz, S. (1999). High rates of conjugation in bacterial biofilms as determined by quantitative *in situ* analysis. *Appl. Environ. Microbiol.* 65, 3710–3713
- Hoiby, N., Bjarnsholt, T., Givskov, M., Molin, S., and Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *Int. J. Antimicrob. Agents* 35, 322–332. doi: 10.1016/j.ijantimicag.2009.12.011
- Huebner, J., and Goldmann, D. A. (1999). Coagulase-negative staphylococci: role as pathogens. *Annu. Rev. Med.* 50, 223–236. doi: 10.1146/annurev.med.50.1.223
- Khan, S. A., and Novick, R. P. (1983). Complete nucleotide sequence of pT181, a tetracycline resistance plasmid from *Staphylococcus aureus*. *Plasmid* 10, 251–259. doi: 10.1016/0147-619X(83)90039-2
- Khan, S. A., Carleton, S. M., and Novick, R. P. (1981). Replication of plasmid pT181 DNA *in vitro*: requirement for a plasmid-encoded product. *Proc. Natl. Acad. Sci. U.S.A.* 78, 4902–4906. doi: 10.1073/pnas.78.8.4902
- Kloos, W. E., and Bannerman, T. L. (1994). Update on clinical significance of coagulase-negative staphylococci. *Clin. Microbiol. Rev.* 7, 117–140. doi: 10.1128/CMR.7.1.117
- Le, K. Y., Dastgheyb, S., Ho, T. V., and Otto, M. (2014). Molecular determinants of staphylococcal biofilm dispersal and structuring. *Front. Cell. Infect. Microbiol.* 4:167. doi: 10.3389/fcimb.2014.00167
- Mack, D., Becker, P., Chatterjee, I., Dobinsky, S., Knobloch, J. K., Peters, G., et al. (2004). Mechanisms of biofilm formation in *Staphylococcus epidermidis* and *Staphylococcus aureus*: functional molecules, regulatory circuits, and adaptive responses. *Int. J. Med. Microbiol.* 294, 203–212. doi: 10.1016/j.ijmm.2004.06.015
- Martineau, F., Picard, F. J., Lansac, N., Menard, C., Roy, P. H., Ouellette, M., et al. (2000). Correlation between the resistance genotype determined by multiplex PCR assays and the antibiotic susceptibility patterns of *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Antimicrob. Agents Chemother.* 44, 231–238. doi: 10.1128/AAC.44.2.231-238.2000
- Miele, A., Bandera, M., and Goldstein, B. P. (1995). Use of primers selective for vancomycin resistance genes to determine van genotype in enterococci and to study gene organization in VanA isolates. *Antimicrob. Agents Chemother.* 39, 1772–1778. doi: 10.1128/AAC.39.8.1772
- Molin, S., and Tolker-Nielsen, T. (2003). Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Curr. Opin. Biotechnol.* 14, 255–261. doi: 10.1016/S0958-1669(03)00036-3
- Monod, M., Mohan, S., and Dubnau, D. (1987). Cloning and analysis of ermG, a new macrolide-lincosamide-streptogramin B resistance element from *Bacillus sphaericus*. *J. Bacteriol.* 169, 340–350. doi: 10.1128/jb.169.1.340-350.1987
- Novick, R. P. (1989). Staphylococcal plasmids and their replication. *Annu. Rev. Microbiol.* 43, 537–565. doi: 10.1146/annurev.mi.43.100189.002541
- Otto, M. (2008). Staphylococcal biofilms. *Curr. Top. Microbiol. Immunol.* 322, 207–228. doi: 10.1007/978-3-540-75418-3\_10
- Otto, M. (2012). Molecular basis of *Staphylococcus epidermidis* infections. *Semin. Immunopathol.* 34, 201–214. doi: 10.1007/s00281-011-0296-2
- Palmer, A. C., and Kishony, R. (2013). Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat. Rev. Genet.* 14, 243–248. doi: 10.1038/nrg3351
- Parra-Ruiz, J., Vidallac, C., and Rybak, M. J. (2012). Macrolides and staphylococcal biofilms. *Rev. Esp. Quimioter.* 25, 10–16.
- Paulsen, I. T., Banerjee, L., Myers, G. S., Nelson, K. E., Seshadri, R., Read, T. D., et al. (2003). Role of mobile DNA in the evolution of vancomycin resistant *Enterococcus faecalis*. *Science* 299, 2071–2074. doi: 10.1126/science.1080613
- Perreten, V., Vorlet-Fawer, L., Slickers, P., Ehrlich, R., Kuhnert, P., and Frey, J. (2005). Microarray-based detection of 90 antibiotic resistance genes of gram-positive bacteria. *J. Clin. Microbiol.* 43, 2291–2302. doi: 10.1128/JCM.43.5.2291-2302.2005

- Pfaller, M. A., and Herwaldt, L. A. (1988). Laboratory, clinical and epidemiological aspects of coagulase-negative staphylococci. *Clin. Microbiol. Rev.* 1, 281–299. doi: 10.1128/CMR.1.3.281
- Reisner, A., Höller, B. M., Molin, S., and Zechner, E. L. (2006). Synergistic effects in mixed *Escherichia coli* biofilms: conjugative plasmid transfer drives biofilm expansion. *J. Bacteriol.* 188, 3582–3588. doi: 10.1128/JB.188.10.3582-3588.2006
- Sambrook, J., and Russel, D. W. (2001). *Molecular Cloning. A Laboratory Manual*. New York, NY: Cold Spring Harbor Laboratory Press.
- Schiwon, K., Arends, K., Rogowski, K. M., Fürch, S., Prescha, K., Sakinc, T., et al. (2013). Comparison of antibiotic resistance, biofilm formation and conjugative transfer of *Staphylococcus* and *Enterococcus* isolates from International Space Station and Antarctic Research Station Concordia. *Microb. Ecol.* 65, 638–651. doi: 10.1007/s00248-013-0193-4
- Schwarz, F. V., Perreten, V., and Teuber, M. (2001). Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from *Enterococcus faecalis* RE25. *Plasmid* 46, 170–187. doi: 10.1006/plas.2001.1544
- Shearer, J. E., Wireman, J., Hostetler, J., Forberger, H., Borman, J., Gill, J., et al. (2011). Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. *G3* 1, 581–591. doi: 10.1534/g3.111.000760
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 3434–3452. doi: 10.1128/MMBR.00020-10
- Stewart, P. S., and Costerton, J. W. (2001). Antibiotic resistance of bacteria in biofilms. *Lancet* 358, 135–138. doi: 10.1016/S0140-6736(01)05321-1
- Tenover, F. C., and Rasheed, J. K. (2004). “Detection of antimicrobial resistance genes and mutations associated with antimicrobial resistance in microorganisms,” in *Molecular Microbiology: Diagnostics Principles and Practice, Vol. 1*, eds D. H. Persing, F. G. Tenover, J. Versalovic, T. Yuer, W. T. J. Relman (Washington, DC: ASM Press), 391–406.
- Tong, S. Y., Chen, L. F., and Fowler, V. G. Jr. (2012). Colonization, pathogenicity, host susceptibility, and therapeutics for *Staphylococcus aureus*: what is the clinical relevance? *Semin. Immunopathol.* 34, 185–200. doi: 10.1007/s00281-011-0300-x
- van Hoek, A. H., Mevius, D., Guerra, B., Mullany, P., Roberts, A. P., and Aarts, H. J. (2011). Acquired antibiotic resistance: an overview. *Front. Microbiol.* 2:203. doi: 10.3389/fmicb.2011.00203
- Wilkins, B. M., and Frost, L. S. (2001). “Mechanisms of gene exchange between bacteria,” in *Molecular Medical Microbiology*, ed M. Sussman (London: Academic Press), 355–400.
- Yang, X., Ma, Q., and Wood, T. K. (2008). The R1 conjugative plasmid increases *Escherichia coli* biofilm formation through an envelope stress response. *Appl. Environ. Microbiol.* 74, 2690–2699. doi: 10.1128/AEM.02809-07
- Zhao, Y. L., Zhou, Y. H., Chen, J. Q., Huang, Q. Y., Han, Q., Liu, B., et al. (2015). Quantitative proteomic analysis of sub-MIC erythromycin inhibiting biofilm formation of *S. suis* in vitro. *J. Proteomics* 116, 1–14. doi: 10.1016/j.jprot.2014.12.019
- Zmantar, T., Kouidhi, B., Miladi, H., and Bakhrouf, A. (2011). Detection of macrolide and disinfectant resistance genes in clinical *Staphylococcus aureus* and coagulase-negative staphylococci. *BMC Res. Notes* 4:453. doi: 10.1186/1756-0500-4-453

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer GDS and handling Editor declared their shared affiliation.

Copyright © 2017 Águila-Arcos, Álvarez-Rodríguez, Garaiyurrebaso, Garbisu, Grohmann and Alkorta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# PCR-Based Analysis of ColE1 Plasmids in Clinical Isolates and Metagenomic Samples Reveals Their Importance as Gene Capture Platforms

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Fabián Lorenzo,  
Universidad de La Laguna, Spain  
Raul Fernandez-Lopez,  
University of Cantabria, Spain  
Ellen Lorraine Zechner,  
University of Graz, Austria  
Antonio Juárez,  
Universitat de Barcelona, Spain

### \*Correspondence:

Bruno Gonzalez-Zorn  
bgzorn@ucm.es

### † Present Address:

Alfonso Santos-Lopez,  
Department of Microbiology and  
Molecular Genetics, University of  
Pittsburgh, Pittsburgh, PA,  
United States

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 15 November 2017

**Accepted:** 28 February 2018

**Published:** 16 March 2018

### Citation:

Ares-Arroyo M, Bernabe-Balas C,  
Santos-Lopez A, Baquero MR,  
Prasad KN, Cid D, Martin-Espada C,  
San Millan A and Gonzalez-Zorn B  
(2018) PCR-Based Analysis of ColE1  
Plasmids in Clinical Isolates and  
Metagenomic Samples Reveals Their  
Importance as Gene Capture  
Platforms. *Front. Microbiol.* 9:469.  
doi: 10.3389/fmicb.2018.00469

**Manuel Ares-Arroyo<sup>1</sup>, Cristina Bernabe-Balas<sup>1</sup>, Alfonso Santos-Lopez<sup>1†</sup>,  
Maria R. Baquero<sup>2</sup>, Kashi N. Prasad<sup>3</sup>, Dolores Cid<sup>1</sup>, Carmen Martin-Espada<sup>1</sup>,  
Alvaro San Millan<sup>4</sup> and Bruno Gonzalez-Zorn<sup>1\*</sup>**

<sup>1</sup> Departamento de Sanidad Animal and Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Facultad de Veterinaria, Universidad Complutense de Madrid, Madrid, Spain, <sup>2</sup> Departamento de Microbiología, Facultad de Veterinaria, Universidad Alfonso X el Sabio, Madrid, Spain, <sup>3</sup> Department of Microbiology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India, <sup>4</sup> Servicio de Microbiología Hospital Universitario Ramón y Cajal, Instituto de Investigación Sanitaria (IRYCIS), Madrid, Spain

ColE1 plasmids are important vehicles for the spread of antibiotic resistance in the Enterobacteriaceae and Pasteurellaceae families of bacteria. Their monitoring is essential, as they harbor important resistant determinants in humans, animals and the environment. In this work, we have analyzed ColE1 replicons using bioinformatic and experimental approaches. First, we carried out a computational study examining the structure of different ColE1 plasmids deposited in databases. Bioinformatic analysis of these ColE1 replicons revealed a mosaic genetic structure consisting of a host-adapted conserved region responsible for the housekeeping functions of the plasmid, and a variable region encoding a wide variety of genes, including multiple antibiotic resistance determinants. From this exhaustive computational analysis we developed a new PCR-based technique, targeting a specific sequence in the conserved region, for the screening, capture and sequencing of these small plasmids, either specific for Enterobacteriaceae or specific for Pasteurellaceae. To validate this PCR-based system, we tested various collections of isolates from both bacterial families, finding that ColE1 replicons were not only highly prevalent in antibiotic-resistant isolates, but also present in susceptible bacteria. In Pasteurellaceae, ColE1 plasmids carried almost exclusively antibiotic resistance genes. In Enterobacteriaceae, these plasmids encoded a large range of traits, including not only antibiotic resistance determinants, but also a wide variety of genes, showing the huge genetic plasticity of these small replicons. Finally, we also used a metagenomic approach in order to validate this technique, performing this PCR system using total DNA extractions from fecal samples from poultry, turkeys, pigs and humans. Using Illumina sequencing of the PCR products we identified a great diversity of genes encoded by ColE1 replicons, including different antibiotic resistance determinants, supporting the previous results achieved with the collections of bacterial isolates. In addition, we detected cryptic ColE1 plasmids in both families with no known genes in



their variable region, which we have named sentinel plasmids. In conclusion, in this work we present a useful genetic tool for the detection and analysis of ColE1 plasmids, and confirm their important role in the dissemination of antibiotic resistance, especially in the Pasteurellaceae family of bacteria.

**Keywords:** antibiotic resistance, ColE1 plasmids, detection PCR, capture PCR, sentinel plasmids

## INTRODUCTION

Plasmids are autonomously replicating fragments of extra-chromosomal DNA that can be transferred horizontally between bacteria. They usually harbor genes that confer a selective advantage under adverse conditions, becoming a major source of genetic variability in bacteria and playing a key role in their adaptation and evolution (Baquero, 2011; Wiedenbeck and Cohan, 2011). Antimicrobial resistance has become one of the most serious problems in public health, and the concern about the ability of plasmids to spread antimicrobial resistance determinants has greatly increased (Carattoli, 2013).

ColE1-type plasmids (hereafter ColE1 plasmids) are small, mobilizable, multi-copy replicons, found mainly in Enterobacteriaceae and Pasteurellaceae (Tomizawa et al., 1977; San Millan et al., 2007), although they have been described in other families of bacteria (Pan et al., 2010; Vincent et al., 2016). These plasmids have a distinctive theta replication mechanism, regulated by two small RNAs encoded close to the origin of replication (*oriV*) (Tomizawa, 1984; Lilly and Camps, 2015). According to the sequences of their mobilization genes, ColE1 replicons belong to the MOB<sub>p</sub> family of plasmids (Garcillán-Barcia et al., 2009, 2015).

Several works have shown that ColE1 plasmids are carriers of resistance mechanisms of high clinical relevance. In Enterobacteriaceae, these plasmids are present in commensal (Pallecchi et al., 2010, 2011; Anantham and Hall, 2012; Moran and Hall, 2017) and pathogenic isolates (de Toro et al., 2013; Garbari et al., 2015; Stoesser et al., 2016, 2017; Albornoz et al., 2017), where they have been found to carry genes conferring resistance to fluoroquinolones, aminoglycosides, sulfonamides and  $\beta$ -lactams, including antibiotics of last resort such as carbapenems (Papagiannitsis et al., 2015) and even colistin (Borowiak et al., 2017). In Pasteurellaceae, several ColE1 plasmids conferring resistance to tetracyclines, aminoglycosides, sulfonamides and  $\beta$ -lactams have been described in human and animal pathogens of the genera *Pasteurella* (San Millan et al., 2009), *Haemophilus* (San Millan et al., 2010, 2011; Tristram et al., 2010; Moleres et al., 2015) and *Actinobacillus* (Blanco et al., 2007). Furthermore, in Pasteurellaceae species of animal origin, ColE1 plasmids are probably the main vehicle for the acquisition of antibiotic resistance determinants (Lancashire et al., 2005; Blanco et al., 2006, 2007; San Millan et al., 2007), being postulated as key strategy for multidrug resistance in this family (San Millan et al., 2009).

In this study, we carried out a computational analysis of the genetic structure of ColE1 plasmids in Enterobacteriaceae and Pasteurellaceae. With this data, we developed a novel PCR-based strategy for the detection, capture and study of ColE1 plasmids, and validated this technique using antibiotic susceptible and

resistant isolates from Enterobacteriaceae and Pasteurellaceae. The results presented here highlight the importance of ColE1 plasmids as a source of antibiotic resistance, providing a useful tool for the study of these small replicons.

## MATERIALS AND METHODS

### Bacterial Strains, Sample Collection and Culture Conditions

In this study we used a total number of 135 Pasteurellaceae and 50 Enterobacteriaceae isolates. A collection of 44 *Pasteurella multocida* and 39 *Mannheimia haemolytica* isolates were obtained from the lungs of 3-month-old lambs and are described in Supplementary Table 1. These strains were isolated, identified on the basis of phenotype, this identification being confirmed by species-specific PCR, and then further characterized by serotyping (Fraser et al., 1983; Townsend et al., 1998, 2001; Angen et al., 2002). In addition, 52 Pasteurellaceae isolates from dogs and cats were recovered during 2009–2010, from oral samples cultured on Columbia 5% sheep blood agar plus 16 mg/l bacitracin, and characterized by Gram staining, oxidase tests and a lack of growth on MacConkey agar (BioMérieux, France). They were further characterized with the API 20NE microorganism identification test kit (BioMérieux, France) and by species-specific PCR and sequencing of the 16S rRNA gene (Król et al., 2011). *Haemophilus influenzae* and *Haemophilus parasuis* were cultured on chocolate agar PolyViteX plates (BioMérieux, France) and in *Haemophilus* Test Medium (HTM) broth (Wider, Francisco Soria Melguizo, S.A., Spain) at 37°C in microaerophilic conditions (5% CO<sub>2</sub>) for 48 h. Moreover, 50 multidrug-resistant Enterobacteriaceae isolates were obtained from the Sanjay Gandhi Postgraduate Institute of Medical Sciences, in Lucknow, India. All the strains were isolated in 2010 and their antibiotic resistance and plasmid profiles are described in Supplementary Table 2. The remaining Enterobacteriaceae species and Pasteurellaceae species were cultured on Columbia agar +5% sheep blood plates and in BHI broth (BioMérieux, France) at 37°C for 24 h.

Additionally, we collected fecal samples from different origins: poultry, turkey, pig and human. Animal fecal samples were collected from three different farms located in the center area of Spain during 2015–2016. Each farm produced a different animal species: poultry, turkey and pig, respectively. Twenty-five individual samples of feces were taken from each animal species in each farm, and then pooled into one unique sample, stored at –80° until DNA extraction. Additionally, a fecal sample of a human from the same region was also collected during the same period and stored at the same temperature.

## Antibiotic Susceptibility Testing

Antimicrobial susceptibility was determined by disk diffusion and microdilution methods, in accordance with CLSI guidelines (CLSI, 2013a,b). Commercially prepared dehydrated Sensititre panels (Trek Diagnostics, Inc., Westlake, OH) were used for MIC determination. Quality control of the panels was performed according to the manufacturer's instructions. Antibiotic disks were obtained from BioMérieux (BioMérieux, France) and Oxoid (Oxoid Ltd., Basingstoke, United Kingdom). Antibiotics were supplied by Merck (Merck KGaA, Darmstadt, Germany) and Sigma-Aldrich (Sigma Chemical Co. St Louis, Mo, USA).

## Computational DNA Analysis

Data on complete plasmid sequences were obtained from GenBank. The structure of ColE1 from Pasteurellaceae was studied in the complete sequence of 24 wild type plasmids (Figure 1). The structure of ColE1 from Enterobacteriaceae was studied in the complete sequence of 37 wild type plasmids (Figure 2), excluding the numerous sequences from cloning vectors or other ColE1-based genetic tools. For the recovery of wild type ColE1 plasmids, short input sequence BLAST of different 20 bp segments including the origin of replication of ColE1 plasmids was performed (<http://blast.ncbi.nlm.nih.gov/>). The BLASTs were performed excluding all ColE1-derived cloning vectors, and only ColE1 plasmids from wild type strains were analyzed. The conserved region of plasmids was determined independently in Pasteurellaceae and Enterobacteriaceae by sequence alignments using Megablast. Plasmid sequence identity score of plasmid conserved regions was established using ColE1 (Accession no. J01566) and pB1000 (Accession no. DQ840517) as prototypes for ColE1 plasmids from Enterobacteriaceae and Pasteurellaceae families, respectively. Nucleotide BLASTs of the conserved region were performed to establish sequence identity using Nblast and Megablast. Conserved region of plasmid ColE1 was established from nucleotide 641 to 3,940, from the start of RNAII to the end of *mbeE* gene. The conserved region of plasmid pB1000 was established from nucleotide 3,584 to 2,665, excluding *bla<sub>ROB-1</sub>* gene. Highly related sequences were selected to determine the GC content (%) in the conserved region using Serial Cloner 2.1 (Serial Basics, France). The regions with no homology were used to estimate Guanine plus Cytosine percentage in the variable region of the plasmids. Differences between variances were tested with the F test to compare the variances of two samples from normal populations. Different software were used for sequence analysis: 4Peaks 1.6 (Mek&Tosj, Netherlands) for the analysis of chromatograms, NIH online analysis tools (<http://www.ncbi.nlm.nih.gov>) for DNA alignments, and DNA Strider 1.4f13 (CEA, France) and CLC DNA workbench (CLC bio, Cambridge, MA) for general features like full integration of the data or analysis of the annotated sequences.

## In Vitro DNA Analysis

The plasmids used in this study are shown in Figures 1, 2, 4, 6. Plasmid DNA from bacteria isolates was extracted with Plasmid Midi and QIAprep Spin Miniprep (Qiagen, Inc., Chatworth, California, USA). PCR products were purified with Qiagen PCR

Purification or Gel Extraction kits (Qiagen, Inc., Chatworth, California, USA). PCR was performed with Taq polymerase from Biotools (B&M Labs, Spain), Phusion high-fidelity DNA polymerase (Finnzymes, Woburn, MA, USA), AmpliTaq Gold DNA polymerase (Applied Biosystems, AB, Foster City, CA, USA) and Taq-Core (Qbiogene, Carlsbad, CA, USA). Automated Sanger sequencing of six complete plasmids was carried out with an Abi-Prism Apparatus (Perkin-Elmer) at Secugen S. L. (Madrid, Spain). The *P. multocida* isolates were identified by PCR detection of the *kmt* gene with the species-specific primers KMT1T7 and KMT1SP6 (Townsend et al., 1998). Capsular type was determined by the PCR method described by Townsend et al. (2001).

## Detection PCR for ColE1 Plasmids

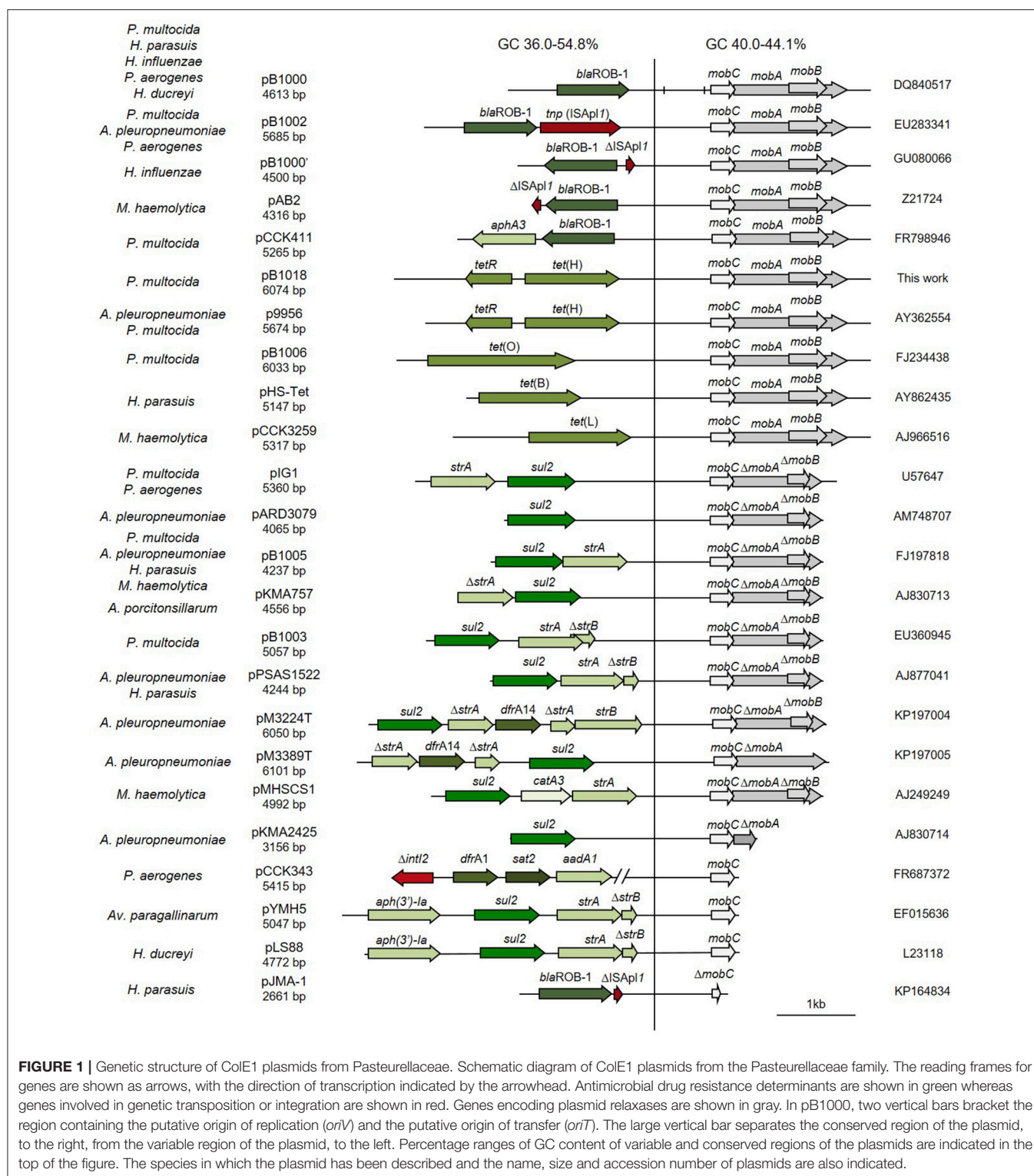
DNA-free polymerases, such as AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA) and Taq-Core (Qbiogene, Carlsbad, CA, USA), are required for this test in Enterobacteriaceae species to avoid false positive results (Supplementary Figure 1). Crude bacterial lysate can act as DNA template for this reaction. Detection PCR was performed according to the kit manufacturer's instructions, with primers ColE1 detF (tgaacgggggttcgtgca)/ColE1 detR (cgttttccataggtccgcc) for Enterobacteriaceae, producing a PCR product of about 300 bp; and ColE1-P detF (gtctccgttcgtgctacggt)/ColE1-P detR (aaatcagcggagccgataggc) for Pasteurellaceae, producing a PCR product of about 450 bp. The amplification conditions were: initial denaturation for 5 min at 94°C, followed by 25 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 55°C and extension for 30 s at 72°C, with a final extension phase for 10 min at 72°C.

## Capture PCR for ColE1 Plasmids

Capture PCR was performed with the Phusion high-fidelity DNA polymerase (Finnzymes, Woburn, MA, USA). No false positive results were observed using this polymerase for the capture PCR. This PCR has been shown to amplify pUC19 vectors with insertions generating products of up to 15 kb (Figure 3). Plasmid preparations using kits as the QIAprep Spin Miniprep (Qiagen, Inc., Chatworth, California, USA) are the recommended DNA template. PCR was performed with the ColE1 cap-1 (tgcacgaacccccgttca)/ColE1 cap-2 (ggcggagcctatggaaaaacg) primers for Enterobacteriaceae and the ColE1-P cap-1 (accgtagcagaaacggagac)/ColE1-P cap-2 (gcctatcggtccgctgattt) primers for Pasteurellaceae, according to the kit manufacturer's instructions. The following conditions were used: initial denaturation for 30 s at 98°C, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 56°C for 10 s and extension at 72°C for 4 min, with a final extension phase for 10 min at 72°C.

## Sequencing and Analysis of Metagenomic Samples

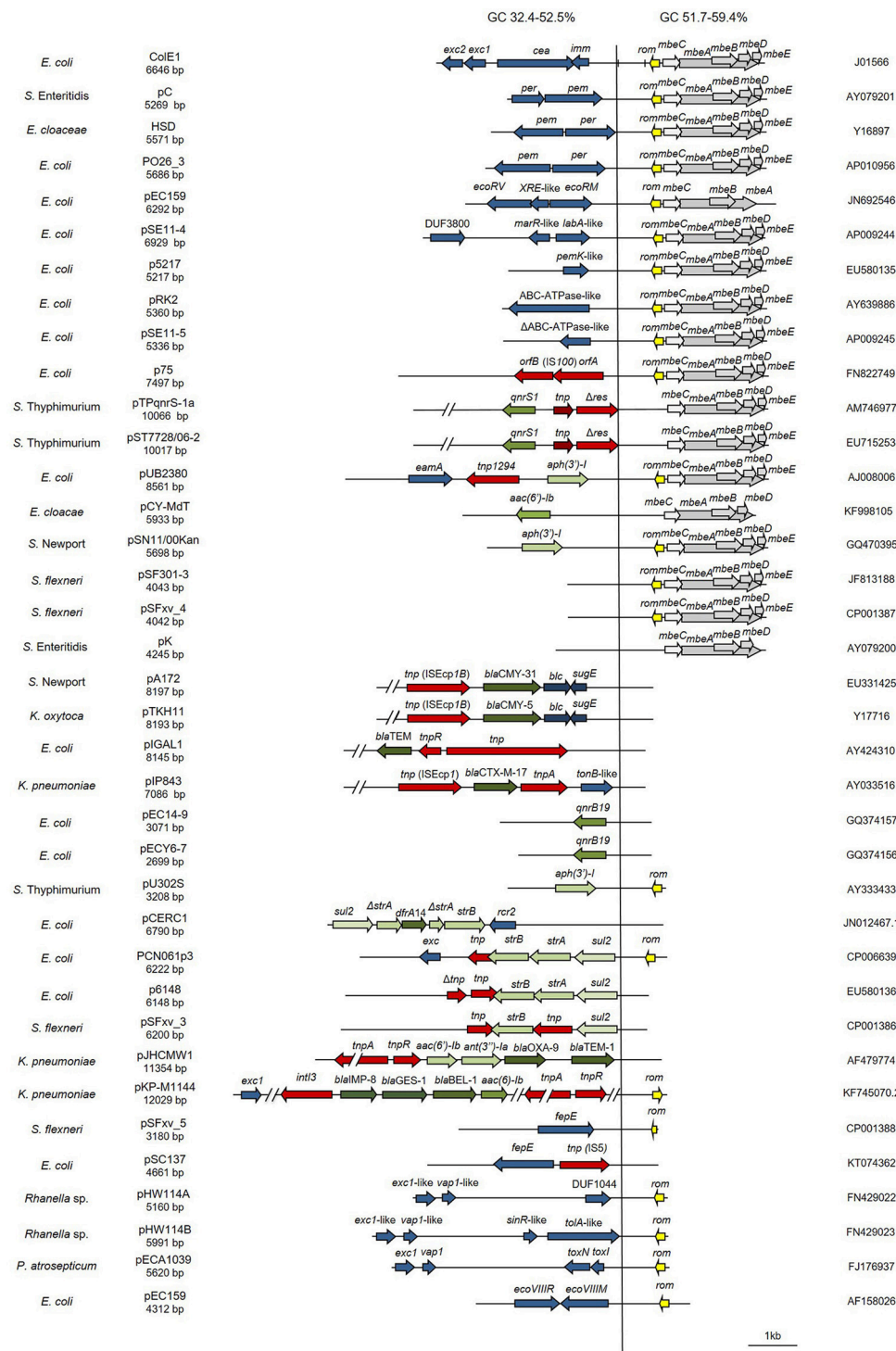
We extracted the total DNA from the four fecal samples with QIAmp Fast DNA Stool Mini Kit (Qiagen, Inc., Chatworth, California, USA), and then performed the Capture PCR from this



total DNA. After purifying the amplified DNA with the Qiagen PCR Purification kit (Qiagen, Inc., Chatworth, California, USA), it was sequenced and assembled at MicrobesNG (Birmingham, United Kingdom) by Next Generation Sequencing following their standard analysis pipeline.

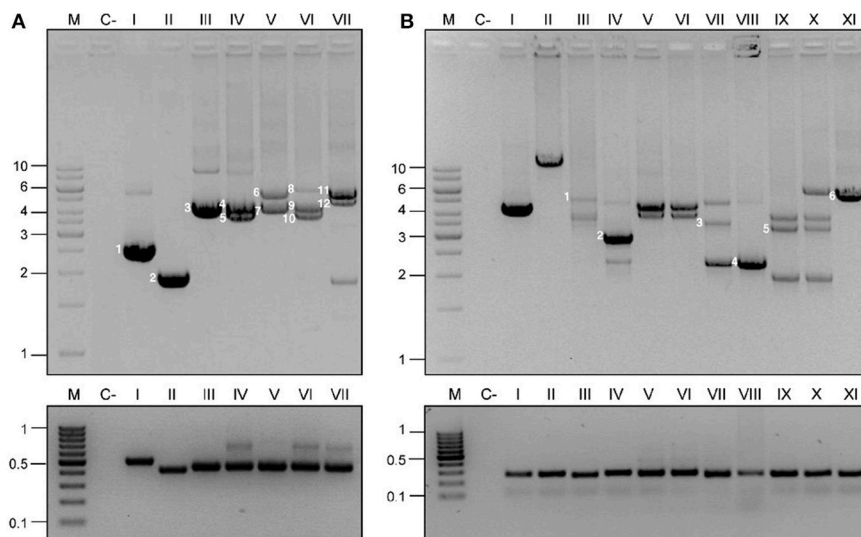
DNA was quantified in triplicates with the Quantit dsDNA HS assay in an Eppendorf AF2200 plate reader. Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with two modifications: two nanograms of DNA instead





**FIGURE 2 |** Genetic structure of ColE1 plasmids from Enterobacteriaceae. Schematic diagram of the 37 ColE1 plasmids from the Enterobacteriaceae family studied in this work. The reading frames for genes are shown as arrows, with the direction of transcription indicated by the arrowhead. The names of the genes, or the names of the family of proteins they encode, are indicated. Antimicrobial drug resistance genes are shown in green and genes involved in genetic transposition or integration are shown in red. Genes encoding plasmid relaxases are shown in gray and the *rom* gene implicated in the regulation of plasmid replication is shown in yellow. The remaining ORFs are shown in blue. In ColE1, two vertical bars bracket the region containing the origin of replication (*oriV*) and the origin of transfer (*oriT*). The large vertical bar separates the conserved region of the plasmids, to the right, from the variable region of the plasmids, to the left. Percentage ranges of GC content of variable (**Left**) and conserved (**Right**) regions of the plasmids are indicated in the top of the figure. The species in which the plasmid has been described, and the name, size, and accession number of plasmids are also indicated.





**FIGURE 3 |** Detection and capture PCR for ColE1 plasmids. ColE1 detection (bottom) and capture (top) PCRs in Pasteurellaceae (**A**) and Enterobacteriaceae (**B**). M stands for molecular weight marker and numbers to the left of the panels indicate the size in kb. C- Stands for negative control. The lanes correspond to different strains and are indicated with roman numbers. (**A**) Pasteurellaceae family PCRs. Negative control (c-) corresponds to *H. influenzae* RdKW20. The products of the capture PCR, corresponding to the ColE1 plasmids carried by the strains, are indicated by numbers. Lane I, *P. stomatis* BB1086: pB000a (1). Lane II, *Frederiksenia canicola* BB1087: pB000b (2). Lane III, *H. influenzae* BB1059: pB1000 (3). Lane IV, *P. multocida* BB1035: pB1000 (4) and pB1005 (5). Lane V, *P. multocida* BB1041: p9956 (6) and pB1000 (7). Lane VI, *P. multocida* BB1044: pB1000 (9), pB1005 (10) and pB1006 (8). Lane VII, *P. multocida* BB1046: pB1002 (11) and pB1003 (12). (**B**) Enterobacteriaceae family PCRs. Negative control (c-) represents *E. coli* DH5 $\alpha$ . Lanes I and II correspond to ColE1 based cloning vectors pTOPO and pUC19 (with the insertion of a ~13 kb DNA fragment), respectively. Lanes III to XI correspond to wild type strains from the Sanjay Gandhi Postgraduate Institute of Medical Sciences in India. Lanes III-VII, *K. pneumoniae*. Lane VIII, *P. mirabilis*. Lanes IX and X, *E. cloacae*. Lane XI, *E. coli*. Six random plasmids were completely sequenced from these strains, and are indicated by numbers in the agarose gel: pB1019 (1), pB1020 (2), pB1022 (4), pB1023 (5), and pB1024 (6).

of one were used as input, and PCR elongation time was increased to 1 min from 30 s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the KapaBiosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina MiSeq using a 250 bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cutoff of Q15 (Bolger et al., 2014). *De novo* assembly was performed on samples using SPAdes version 3.7 (Bankevich et al., 2012). Additional information about the data of this sequencing is presented in Supplementary Table 3.

An automated annotation of the contigs assembled was performed at MicrobesNG using Prokka (Seemann, 2014). Moreover, we developed an additional annotation of the sequences combining the analysis tools RAST (Aziz et al., 2008) and ResFinder (Zankari et al., 2012). In order to assure that the annotations were actually present in ColE1 replicons, we analyzed the genetic environment of the genes in the contigs assembled at MicrobesNG. We assumed that these genes were encoded actually on ColE1 replicons when their whole contig sequence upstream and downstream the gene corresponded to ColE1-like sequences according to the NIH online analysis tool Nucleotide BLAST (<http://blast.ncbi.nlm.nih.gov/>).

## Nucleotide Sequence Accession Numbers

Nucleotide sequences of the plasmids obtained in this study have been deposited in GenBank under the following accession

numbers: pB1018 from BB1253, JQ319774; pB000a from BB1086, JQ319773; pB000b from BB1087, JQ319771; pB1019 from BB1088, JQ319775; pB1020 from BB1089, JQ319772; pB1021 from BB1090, JQ319767; pB1022 from BB1091, JQ319766; pB1023 from BB1092, JQ319770; and pB1024 from BB1093, JQ319768.

## RESULTS

### Computational Analysis of ColE1 Plasmids ColE1 Plasmids in Pasteurellaceae

The bioinformatic analysis of the ColE1 plasmids from Pasteurellaceae species led to the identification of two differentiated genetic regions (**Figure 1**): a conserved region carrying all the elements controlling replication and transfer, and a variable region encoding mainly antibiotic resistance genes. The conserved region of ColE1 plasmids from Pasteurellaceae had an average size of 2,513 bp (Standard Deviation,  $SD = 718$  bp) and was highly similar among all of the plasmids, with a GC content between 40.0 and 44.1% and high nucleotide sequence identity (average = 97.62%,  $SD = 2.06\%$ ) (**Figure 1**). In contrast, the variable region of these ColE1 plasmids presented a high level of genetic divergence. As result of this variability, the GC content in this region varied between 36.0 and 54.8%. The variance of the GC content was significantly lower in the conserved region than in the variable region [ $F_{(23)} = 40.22$ ,  $P < 0.001$ ].

## ColE1 Plasmids in Enterobacteriaceae

In this family, the ColE1 replicons also showed a conserved area involved in plasmid housekeeping functions and a variable region. However, in contrast to plasmids from Pasteurellaceae, ColE1 plasmids in Enterobacteriaceae encoded a wide variety of accessory genes, including also antibiotic resistance determinants (Figure 2). The conserved region of ColE1 plasmids from Enterobacteriaceae had an average size of 1,817 bp ( $SD = 1,197$  bp) with an average nucleotide sequence identity of 82.59% ( $SD = 10.79\%$ ). This region presented a GC content between 51.7 and 59.4%, whereas in the variable region GC content varied between 32.4 and 52.5%. Again, the variance of the GC content was significantly lower in the conserved region [ $F_{(36)} = 14.11$ ,  $P < 0.001$ ].

## Development of a PCR-Based System for Detection and Capture of ColE1 Plasmids

We used the data obtained from the *in silico* analysis of the sequences of wild type ColE1 plasmids to develop a two PCR-based system for the detection and capture of ColE1 replicons in Enterobacteriaceae and Pasteurellaceae. Pairwise and multiple DNA alignments of the conserved sequences of the plasmids were performed in order to detect conserved regions suitable for the design of PCR primers (for detailed description of the PCRs conditions and primers see Material and Methods).

As the nucleotide sequence of ColE1 plasmids presents a highly conserved region among all the replicons belonging to the same bacterial family (82.59% in Enterobacteriaceae and 97.62% in Pasteurellaceae), but drastically different when compared against plasmids from the other family, we were forced to design two different set of primers, each one specific for Enterobacteriaceae ColE1 plasmids and Pasteurellaceae ColE1 plasmids, respectively.

Thus, we developed a two PCRs system for the specific analysis of ColE1 plasmids in each family of bacteria. First, a “Detection PCR” using a pair of universal primers was designed to amplify a small fragment from the conserved region of the replicons, close to the *oriV*. Second, we designed a “Capture PCR,” using a pair of primers annealing to the exact same region as the primers form the detection PCR, but amplifying outwards, allowing the capture of the variable region of the plasmid. Using this technique, the whole plasmid sequence is available for further analysis.

## Validation of the PCR-Based System for ColE1 Analysis

### Validation in ColE1 Plasmids From Pasteurellaceae

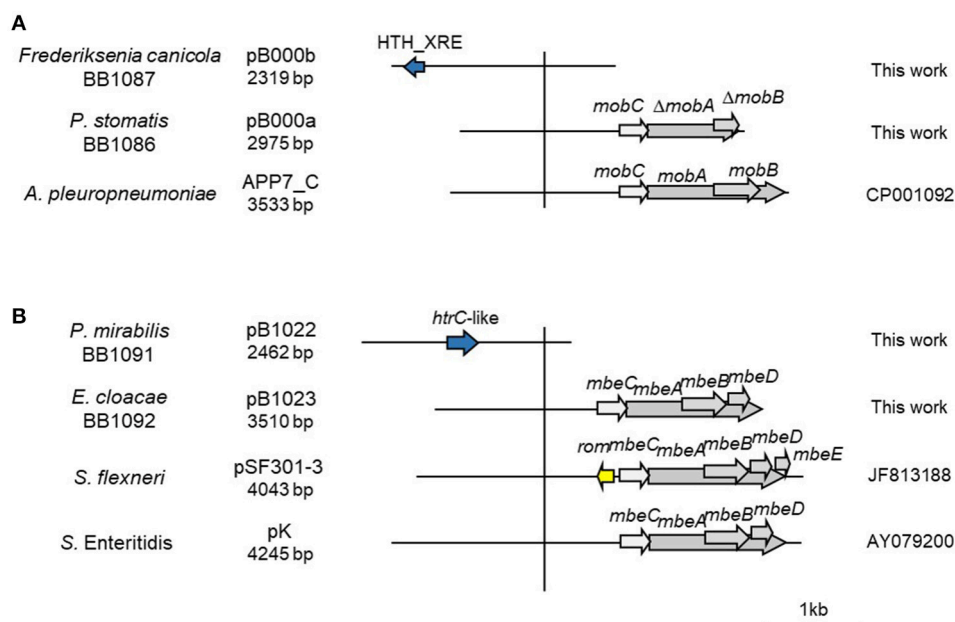
In order to validate the PCR-based system in Pasteurellaceae, we used a well-characterized series of strains of *H. influenzae* (San Millan et al., 2010, 2011), *H. parasuis* (San Millan et al., 2007) and *P. multocida* (San Millan et al., 2009; Santos-Lopez et al., 2017). These strains, previously described by our group, carried one, two or three ColE1 plasmids. The Detection PCR was positive in every isolate and, most importantly, the Capture PCR was not only able to amplify a single ColE1 plasmid, but it gave rise to different PCR products, corresponding to

each coexisting plasmid in those strains carrying multiple (up to three) plasmids. As negative controls we used the reference strains *H. influenzae* RdKW20, *P. multocida* ATCC 43137 and *H. parasuis* ATCC 19417, which do not carry ColE1 plasmids. As expected, no PCR product was observed in any of these strains. In Figure 3A we show the results of a representative group of Pasteurellaceae strains tested: *H. influenzae* RdKW20, *P. stomatis* BB1086, *Frederiksenia canicola* BB1087, *H. influenzae* BB1059, *P. multocida* BB1035, *P. multocida* BB1041, *P. multocida* BB1044, *P. multocida* BB1046.

Additionally, we decided to establish two new collections of Pasteurellaceae isolates to test the PCR-based system in a wider range of species. We constructed a first collection of 52 Pasteurellaceae isolates from oral samples collected from healthy dogs and cats, including strains from six different Pasteurellaceae species: *P. multocida*, *Pasteurella canis*, *Pasteurella pneumotropica*, *Pasteurella stomatis*, *Pasteurella dagmatis* and *F. canicola*. A second collection was collected from lungs of 3-month-old lambs counting with 44 *P. multocida* and 39 *M. haemolytica* strains (Supplementary Table 1). We performed antibiotic susceptibility testing for clinically relevant antibiotics in both collections, detecting eight tetracycline resistant *P. multocida* strains in the lamb collection (Supplementary Table 1). We carried out the ColE1 Detection PCR and detected 10 positive strains; the eight tetracycline resistant *P. multocida* isolates from the lamb collection and two susceptible isolates from the dog collection: a *P. stomatis* (BB1086) and a *F. canicola* (BB1087). The Capture PCR was also positive in these 10 strains. In the eight tetracycline resistant *P. multocida* the Capture PCR revealed the presence of a plasmid of about 6 kb in size in all the isolates. This plasmid, bearing the tetracycline resistance gene *tet(H)*, was completely sequenced and named pB1018 (Figure 1). The strains from the dog collection harbored two plasmids of 2,975 bp (BB1086) and a 2,319 bp (BB1087) in size, respectively. These cryptic plasmids carried no detectable known gene in their variable region, and we named them pB000a and pB000b (Figure 4A).

### Validation in ColE1 Plasmids From Enterobacteriaceae

We first tested the PCR system using ColE1-based cloning vectors, such as pTOPO and pUC19. The Detection PCR was positive and the Capture PCR was able to amplify fragments of up to 15 kb in size from genetic constructions using the ColE1-based pUC19 plasmid (Figure 3B). Therefore, this reaction should be able to capture any ColE1 plasmid from wild type strains. As negative controls we used laboratorial strains carrying no ColE1 plasmids as *Escherichia coli* DH5 $\alpha$ . In contrast to the case of Pasteurellaceae, we did not have access to a previously characterized collection of Enterobacteriaceae strains carrying ColE1 plasmids. Hence, we decided to analyse a new collection of 50 clinical isolates of Enterobacteriaceae, including six different species and displaying resistance to various antibiotics, recovered at the Sanjay Gandhi Postgraduate Institute of Medical Sciences in India (Supplementary Table 2). Thirty seven of fifty isolates gave positive results in the ColE1 Detection PCR (Figure 3B). The amplicons from a representative number of strains were



**FIGURE 4 |** Genetic structure of a selection of ColE1 cryptic plasmids. Schematic diagram of ColE1 cryptic plasmids from Pasteurellaceae (**A**) and Enterobacteriaceae (**B**). The reading frames for genes are shown as arrows, with the direction of transcription indicated by the arrowhead. Genes encoding plasmid relaxases are shown in gray. The vertical bar separates the conserved region of the plasmids, to the right, from the variable region of the plasmids, to the left. The species in which the plasmid has been described, and the name, size, and accession number of plasmids are also indicated.

sequenced, and were confirmed to have originated from ColE1 plasmids. Thus, 74% of the Enterobacteriaceae studied isolates actually carried at least one ColE1 plasmid. The Capture PCR was then performed for the isolates bearing ColE1 plasmids, and it generated from one to three amplicons per isolate, with sizes ranging from 2 to 10 kb, corresponding to the various ColE1 plasmids present in the cell (**Figure 3B**). This result was confirmed by partial sequencing of the different PCR products after DNA purification from agarose gel (see Material and Methods).

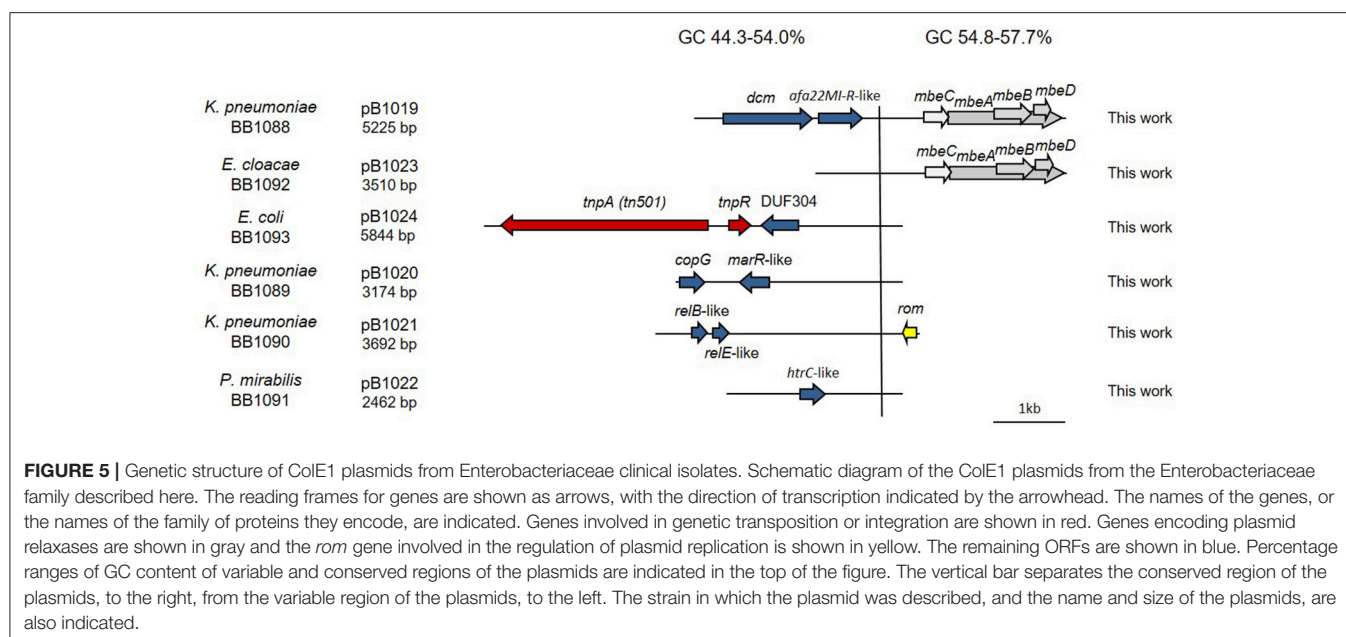
We completely sequenced six random ColE1 plasmids (pB1019 to pB1024) from this collection of strains to confirm the results and the utility of the method developed here (**Figure 5**). These sequenced replicons had the typical characteristics of ColE1 plasmids, with both the conserved and the variable regions. Again, the GC content of the conserved region was very similar among these plasmids (54.8–57.7%), whereas the plasmid variable region presented diverse GC content (44.3–54.0%). We found different genes in the variable region of these replicons: toxin-antitoxin systems (pB1021), restriction-modification systems (pB1019) or transposases like Tn501 (pB1024). Interestingly, we also found two cryptic plasmids (pB1022 and pB1023). However, no antibiotic resistance gene was found in these plasmids despite the high level of resistance of the strains.

It is important to mention that contamination with DNA from ColE1 cloning vectors in some of the commercial DNA polymerases generated a false-positive reaction in the ColE1 Detection PCR in Enterobacteriaceae (Supplementary Figure 1).

DNA-free polymerases, such as AmpliTaq Gold DNA polymerase (Applied Biosystems) and Taq-Core (Qbiogene), should therefore be used for this PCR reaction. Such contamination, leading to erroneous PCR results, has been described before and is of particular relevance in the case of the *bla*<sub>TEM-1</sub> and *bla*<sub>TEM-116</sub> β-lactamases genes (Koncan et al., 2007; Jacoby and Bush, 2016).

### Validation in ColE1 Plasmids From Intestinal Microbiota

In order to validate if the PCR system is useful for the analysis of ColE1 plasmids from metagenomic samples, we decided to test the Capture PCR directly on total DNA extracted from fecal samples using the primers specific for ColE1 plasmids from Enterobacteriaceae. Four pools of fecal samples were tested in this study, three of them collected from healthy animals of different species: poultry, turkey and pig; and a fourth one from human origin. We did not use the PCR-system for Pasteurellaceae given the limited presence of these bacteria in the gut microbiota of both animals and humans (Roto et al., 2015; Burrough et al., 2017; Gupta et al., 2017). After performing the Capture PCR and purifying the final product, we sequenced the amplified DNA resulting from all the ColE1 plasmids harbored in the Enterobacteriaceae cells present in the fecal samples using Illumina MiSeq. We analyzed the genes present on the amplicons, and their genetic environment, confirming that they were actually present in ColE1 plasmids. We found several mobilization genes, transposases, toxin-antitoxin systems, bacteriocins and restriction-modification



systems (Table 1), in addition to several hypothetical proteins ranging between 32 and 418 amino acids. Interestingly we also observed multiple antibiotic resistance genes conferring resistance to some of the most important antibiotic families such as  $\beta$ -lactams, aminoglycosides and fluoroquinolones (Table 1).

## DISCUSSION

In this work, we analyzed the structure and content of ColE1 plasmids described in Pasteurellaceae and Enterobacteriaceae up to date. Our results showed that although ColE1 plasmids are different in these two families, they presented the same genetic structure. We observed two differentiated regions in the plasmids, one highly conserved and another one highly variable. The conserved region harbored all the housekeeping functions of the plasmid, including the origins of replication (*oriV*) and transfer (*oriT*) and, in most of plasmids analyzed, relaxases genes. The GC content of this conserved region is similar to the GC content of the genomes in which ColE1 plasmids have been found (Figure 6): 37–44% in Pasteurellaceae (Supplementary Table 3) and 50–59% in Enterobacteriaceae (Supplementary Table 4), strongly suggesting a plasmid/host adaptation process. On the other hand, the variable region carries accessory genes from different origins. In this variable region we found a wide variety of antibiotic resistance determinants in both families. In Pasteurellaceae, we found genes conferring resistance to tetracycline,  $\beta$ -lactams, aminoglycosides, sulphonamides, trimethoprim and chloramphenicol (Figure 1). In contrast to the plasmids from Pasteurellaceae, ColE1 plasmids from Enterobacteriaceae carried a wide variety of genes apart from antibiotic resistance determinants: genes involved in resistance to phage infections such as abortive





infection systems (Fineran et al., 2009) and restriction-modification systems (Gregorova et al., 2002), genes involved in ferric transport (Ye et al., 2010) or genes encoding bacteriocins, such as the colicin E1 (Tomizawa et al., 1977). However, despite the higher qualitative diversity in the genes encoded in their variable region, ColE1 plasmids from Enterobacteriaceae carried antibiotic resistance genes, mainly against  $\beta$ -lactams, aminoglycosides and sulphonamides (Figure 2).

Using the data obtained from our computational analysis, we developed a new PCR-based system able to detect and completely capture ColE1 plasmids in Enterobacteriaceae and Pasteurellaceae. This is, to the best of our knowledge, the first system of this nature developed for ColE1 plasmids in Pasteurellaceae. In Enterobacteriaceae, previous PCR-based tests (García-Fernández et al., 2009; Chen et al., 2010; Alvarado et al., 2012) have been described for the detection of ColE1 plasmids.

García-Fernández et al. (2009) designed a set of primers targeting a conserved region in the origin of replication of ColE1 plasmids while looking for plasmids harboring quinolone resistance genes in *Salmonella*. With these primers (Table 2) they successfully detected three different ColE1 replicons, demonstrating the efficacy of this PCR. However, by testing these primers *in silico* against the ColE1 represented in Figure 2, just 14 out of the 37 plasmids carried the complete sequence for primers hybridization. In addition, Chen et al. (2010) developed a PCR-based system for the detection of ColE1 plasmids in *Salmonella*, by using primers targeting a conserved region within the origin of replication and the *rom* gene (Table 2). However, we also tested these primers *in silico* and just 13 out of the 37 replicons did harbor the whole sequence complementary to these oligonucleotides. In parallel to the previous techniques targeting the origin of replication, Alvarado et al. (2012)







**TABLE 1** | List of genes harbored by ColE1 plasmids in the fecal samples.

NAME	DESCRIPTION				
<b>REPLICATION</b>					
<i>rom</i>	RNAI modulator protein	*	*	*	*
<b>CONJUGATION</b>					
<i>mbeA</i>	Mobilization protein	*	*	*	*
<i>mbeB</i> -like	Mobilization protein	*			
<i>mbeC</i>	Mobilization protein	*	*	*	*
<i>mbeD</i>	Mobilization protein	*			
<i>virB4</i>	Type IV secretion system protein		*		*
<b>ANTIBIOTIC RESISTANCE</b>					
<i>aph(3')-IIa</i>	Aminoglycoside phosphotransferase	*	*		*
<i>bla</i> <sub>TEM-116</sub>	β-Lactamase	*	*	*	*
<i>strA</i>	Aminoglycoside phosphotransferase		*	*	
<i>strB</i>	Aminoglycoside phosphotransferase		*		
<i>sul2</i>	Sulfonamide-resistance dihydropteroate synthase		*		
<i>qnrB19</i>	Quinolone resistance protein			*	*
<b>TRANSPOSASES</b>					
IS91	Transposase		*		
IS5	Transposase			*	*
IS3	Transposase				*
Tn3	Transposon				*
<b>TOXIN-ANTITOXIN SYSTEMS</b>					
<i>ccdB</i>	CcdB/CcdA Toxin protein	*			*
<i>higA</i>	HigB/HigA Antitoxin protein	*		*	*
<i>higB</i>	HigB/HigA Toxin protein				*
<i>lsoA</i>	LsoA/LsoB Toxin protein				*
<i>parE</i>	ParE/ParD Toxin protein	*		*	
<i>pemI</i>	PemK/PemI Antitoxin protein				*
<i>pemK</i>	PemK/PemI Toxin protein				*
<b>BACTERIOCINS</b>					
<i>caa</i>	Colicin-A	*	*		
<i>cai</i>	Colicin-A immunity protein		*		
<i>cea</i>	Colicin-E1	*	*		
<i>imm</i>	Colicin-E1 immunity protein				*
<i>lys</i>	Colicin-E1 lysis protein				*
<i>csa</i>	Colicin-S4			*	
<i>csi</i>	Colicin-S4 immunity protein			*	
<i>cta</i>	Colicin-10		*	*	*
<i>kba</i>	Klebicin-B				*
<i>kbi</i>	Klebicin-B immunity protein				*
<b>RESTRICTION-MODIFICATION SYSTEMS</b>					
<i>ecoRV</i>	Restriction endonuclease EcoRV	*	*	*	
<i>ecoRVM</i>	DNA-methyltransferase EcoRV	*			
<i>ecoVIII</i>	Restriction endonuclease EcoVIII			*	
<i>ecoVIII</i>	DNA-methyltransferase EcoVIII			*	
<i>banIR</i>	Restriction endonuclease BanI	*		*	*
<i>banIM</i>	DNA-methyltransferase BanI				*
<i>styD4IR</i>	Restriction endonuclease StyD4I	*			*
<i>styD4IM</i>	DNA-methyltransferase StyD4I	*			*

(Continued)

TABLE 1 | Continued

NAME	DESCRIPTION				
<i>eco29kIR</i>	Restriction endonuclease Eco29kl	*			
<i>eco29kIM</i>	DNA-methyltransferase Eco29kl	*			
<i>eco034IR</i>	Restriction Enzyme Eco034I	*			*
<i>eco034IM</i>	DNA-methyltransferase Eco034I	*			
<i>ydiOR</i>	Restriction endonuclease YdiO		*		
<i>ydiOM</i>	DNA-methyltransferase YdiO		*		
<b>MISCELLANEOUS</b>					
<i>abi-like</i>	Abortive infection bacteriophage protein		*	*	*
<i>chrB</i>	Membrane chromate resistance protein				*
<i>copG</i> family	Ribbon-helix-helix protein	*	*	*	
<i>csp</i>	Cold shock protein	*		*	*
<i>era</i>	GTPase Era			*	
<i>exc1</i>	Entry exclusion protein 1	*	*	*	*
<i>exc2</i>	Entry exclusion protein 2	*			*
<i>escC</i> family	Type III secretion system protein	*		*	*
<i>fepE</i>	Ferric enterobactin transport protein	*	*	*	*
Helix-turn-helix	Helix-turn-helix domain containing protein	*	*	*	*
<i>htrC</i>	Heat shock protein C	*	*	*	*
<i>kdgF</i>	Pectin degradation protein				*
<i>nicB</i>	Nicotine dehydrogenase subunit B	*			
<i>nikA</i>	Nickel ABC transporter substrate-binding protein	*	*	*	*
<i>parA</i>	Plasmid partitioning protein		*		
Phage integrase	Phage integrase family	*			
Phosphatase 2C	Type 2C protein phosphatase			*	*
<i>pspF</i>	Phage shock protein operon transcriptional activator			*	*
Reverse transcriptase motif	RNA-dependant DNA-polymerase	*		*	*
<i>rihC</i>	Ribonucleoside hydrolase	*			
<i>tsr</i>	Methyl-accepting chemotaxis protein		*		*
<i>xre</i>	Transcriptional regulator XRE family protein			*	
<i>ydaM</i>	Diguanylate cyclase, csgD regulator		*		
<b>DIMER RESOLUTION</b>					
<i>xerD</i>	Site-specific tyrosine recombinase	*			*
<i>xerC</i>	Site-specific tyrosine recombinase	*			*

The name and description of the genes is given on the left side of the table, while the asterisks (\*) of the right side indicate the presence of each gene in the different samples, representing from left to the right: poultry, pig, turkey and human.

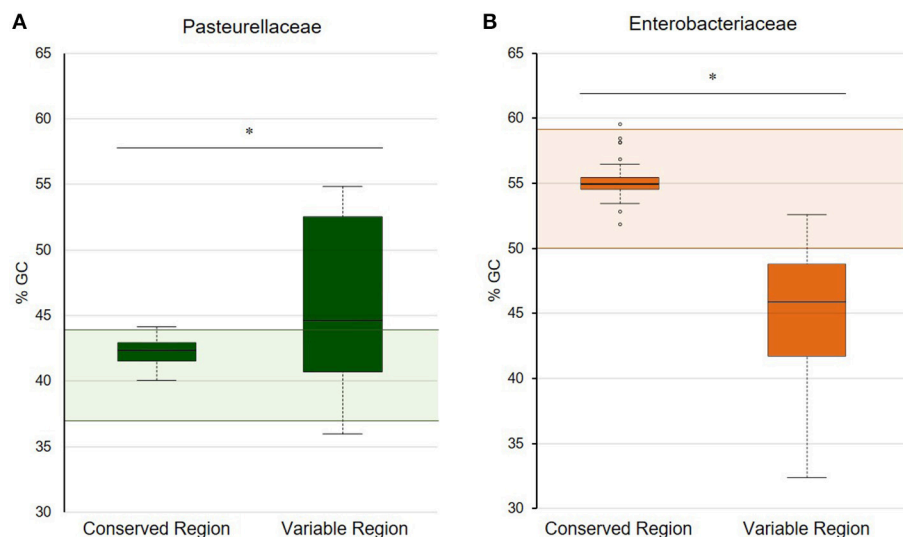
developed a Degenerate Primer MOB Typing (DPMT) technique, extremely useful to detect and classify plasmids present in gamma-proteobacteria by targeting their relaxases genes. This DPMT included different degenerate primers against the MOB<sub>p5</sub> relaxases of ColE1 plasmids (Table 2). However, as these genes are actually absent in a substantial proportion of ColE1 replicons (Figure 2), a considerably part of these plasmids would not be detected by using only this set of primers. In summary, our bioinformatic analysis revealed that these prior methods, although scrupulously designed and useful in the particular studies in which they were employed, would fail to detect part of the wild type ColE1 plasmids described to date, either for lack of sensitivity of the primers or for targeting the mobilization genes. Nevertheless, in order to reach the most sensitivity

as possible, we suggest the combination of all these primers to assure the detection of any ColE1 plasmid present in a sample.

In order to validate our technique, we used the PCR system in a range of bacterial collections from Enterobacteriaceae and Pasteurellaceae as well as in metagenomic samples from fecal origin. These experiments revealed interesting results. In Pasteurellaceae we confirmed the tight link between ColE1 plasmids and antibiotic resistance. Moreover, we discovered the presence of cryptic ColE1 plasmids in antibiotic susceptible *P. stomatis* BB1086 (pB000a) and *F. canicola* BB1087 (pB000b), which only encoded plasmid housekeeping genes. In Enterobacteriaceae, the PCR-based screening system showed that the prevalence of the ColE1 replicons in the antibiotic

**TABLE 2** | Primers used to detect ColE1 plasmids in Enterobacteriaceae isolates in previous studies.

Primer name	Sequence (5'-3')	Length	Reference
oricolE Fw	GTTCGTGCATACAGTCCA	18	García-Fernández et al., 2009
oricolE Rv	GGCGAAACCCGACAGGACT	19	
CC7059F	TTCGTGACACAGCCCA	17	Chen et al., 2010
CC7062R	TGCGGTTATCCACAGAATCA	20	
CC7063F	GCGGACAGGTATCCGGTAA	19	Alvarado et al., 2012
P51-f	TACCACGCCCTATGCGAARAARTAYAC	27	
P52-f	GATAGCCTTGATTTTAATACACCAAYACYTAYAC	35	
P5-r	CCCTGTGCTGGTGYTSNACCCA	23	
P53-f	GGGCTCGCACGAYCAYACNGG	21	This work
P53-r	GCCCAGCCCTTTTCRTGRTTRTG	23	
ColE1 detF	TGAACGGGGGGTTCGTGCA	19	
ColE1 detR	CGTTTTTCCATAGGCTCCGCC	21	



**FIGURE 6** | GC content of the conserved and variable regions of ColE1 plasmids. Representation of the percentage of GC content in both the conserved and variable region of ColE1 plasmids in Pasteurellaceae (A) and Enterobacteriaceae (B). Colored boxplots represent the GC content of each of the ColE1 replicons studied. The shaded areas correspond to the %GC range of the genomes analyzed of these two bacteria families (Supplementary Tables S1, S2). The asterisks indicate the difference of the mean values of GC content between the conserved and variable region in both families of bacteria [ $t_{(24)} = 2.91$ ,  $P < 0.05$  and  $t_{(41)} = 10.83$ ,  $P < 0.05$ , respectively].

resistance collection tested was especially high, with the 74% of the isolates carrying at least one plasmid. We also detected cryptic ColE1 plasmids (pB1022 and pB1023) with no evident genes in their variable region in this collection (Figure 4B). Other cryptic ColE1 plasmids, such as pB000a, pB000b, pB1022, and pB1023, have been previously described in human, animal and environmental isolates (Rozhon et al., 2006; Handford et al., 2009; Bleicher et al., 2013) (Figure 4) and, interestingly, some of the antibiotic resistance genes encoded in ColE1 replicons had been described in different bacterial genus and families (Miranda et al., 2003; Soge et al., 2006; Warburton et al., 2016). Hence, our hypothesis is that these unexpected prevalent cryptic replicons might act as “sentinel plasmids,” capable of maintaining just the conserved region in bacteria due to their capacity of replication

and conjugation (Burian et al., 1997), but able to acquire a wide variety of genes from heterogeneous origins, providing an increased genetic plasticity to their host.

In addition, our metagenomic approach confirmed the large diversity of genes that these small replicons can encode in the gut enterobacteria in healthy animals and humans (Table 1). We consider important to mention that we do not suggest a species distribution of the ColE1 genes based on our sample collection. However, we kept it separated in Table 1, firstly to show that there is no bias in our approach, and secondly, as it could be interesting in future works aiming to study the epidemiology of ColE1 plasmids. Many of the detected genes have well known functions in plasmids biology, as mobilization genes, toxin-antitoxin systems (Moran and Hall, 2017), the previously cited

restriction-modification systems or the bacteriocins. However, the function of other genes found in this sample, and previously described in other ColE1 plasmids such as the *copG*-like genes (de Toro et al., 2013), are still unknown. Of especial relevance in these ColE1 amplicons from fecal samples was the detection of antibiotic resistance determinants against quinolones, aminoglycosides, sulphonamides and  $\beta$ -lactams. Previous works showed that the most prevalent antibiotic resistance genes in human and animal gut microbiomes are those conferring resistance to tetracycline, representing even the 90% of the gut resistome (Durso et al., 2012; Pal et al., 2016). However, although ColE1 plasmids frequently carry tetracycline resistance genes in Pasteurellaceae (Figure 1), they are not common in Enterobacteriaceae (Figure 2) and none has been detected in our approach. Other resistance genes are less represented in animal and human microbiomes, although new techniques with higher sensitivity (Lanza et al., 2018) might highlight the presence of genes that have been underrepresented to date in metagenomics samples. In our study, the genes detected with our ColE1 capture PCR were the *qnrB19* quinolone resistance gene in turkey and human, the *aph(3')-IIa* aminoglycoside phosphotransferases in poultry, pig and human, or the *strA* gene in turkey, which was found both alone and forming a *strA-strB-sul2* complex in the pig sample, being all these genes previously described in ColE1 replicons (Figure 2). The only antibiotic resistance determinant found in all the species was the *bla*<sub>TEM-116</sub> gene, which encodes an extended-spectrum  $\beta$ -lactamase (ESBL) (Lahlaoui et al., 2011). TEM-116 has been recently described as the central node of one of the two TEM clusters that group all the known TEM variants (Zeil et al., 2016), emphasizing its importance in antimicrobial resistance evolution and the role of ColE1 plasmids in its dissemination.

In conclusion, we have developed a simple system to screen and characterize ColE1 plasmids, which will allow monitoring the increasingly relevant role of these plasmids in the spread and evolution of antibiotic resistance in Pasteurellaceae and Enterobacteriaceae.

## AUTHOR CONTRIBUTIONS

MA-A, CB-B, and AS-L have contributed with the design of the work, data collection, analysis and interpretation, drafting the article, and AS-L with critical revision of the draft. MB, CM-E, DC, and KP have contributed with strains and critical

revision of the article for final approval. AS has contributed with the design of the work, data collection and analysis, and critical revision of the draft. BG-Z has contributed with the conception and design of the work, interpretation of the data collection and analysis, critical revision of the draft and final approval of the article.

## FUNDING

This work was supported by grants from the Spanish Ministry of Science and Innovation (BIO 2010-20204, PRI-PIBIN-2011-0915, and BFU2011-14145-E), the European Commission (EC) EVoTAR-282004-FP7, the European Commission (EC) EFFORT-613754-FP7 and the Programa de Vigilancia Sanitaria 2009 AGR/4189 of the Comunidad de Madrid (Spain). MA-A is supported by the Universidad Complutense de Madrid (CT27/16-CT28/16). CB-B is supported by the Spanish Ministry of Education, Culture and Sport (FPU13/06215). DC acknowledges support from the MICINN (AGL2009-10136). MB acknowledges the support from Fundación Universidad Alfonso X el Sabio – Grupo Santander. AS is supported by a Miguel Servet Fellowship from the Instituto de Salud Carlos III (MS15/00012) co-financed by The European Social Fund Investing in your future (ESF).

## ACKNOWLEDGMENTS

We thank Natalia Montero for excellent technical support. We thank R. del Campo and J. L. San Millan for advice and technical assistance in the choice of DNA-free polymerases for ColE1 plasmid detection. We also thank to I. Cuesta for her bioinformatics support and advice with the sequencing data analysis. Finally, we would also like to thank the Veterinary Microbiology Students from the Universidad Alfonso X El Sabio of Madrid (term 2009–2010) for their assistance in the establishment of the collection of Pasteurellaceae isolates from dogs and cats. The authors would like to thank the reviewers for their careful and constructive comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00469/full#supplementary-material>

## REFERENCES

- Albornoz, E., Lucero, C., Romero, G., Quiroga, M. P., Rapoport, M., Guerriero, L., et al. (2017). Prevalence of plasmid-mediated quinolone resistance genes in clinical enterobacteria from Argentina. *Microb. Drug Resist.* 23, 177–187. doi: 10.1089/mdr.2016.0033
- Alvarado, A., Garcillan-Barcia, M. P., and de la Cruz, F. (2012). A degenerate primer MOB typing (DPMT) method to classify gamma-proteobacterial plasmids in clinical and environmental settings. *PLoS ONE* 7:e40438. doi: 10.1371/journal.pone.0040438
- Anantham, S., and Hall, R. M. (2012). pCERC1, a small, globally disseminated plasmid carrying the *dfrA14* cassette in the *strA* gene of the *sul2-strA-strB* gene cluster. *Microb. Drug Resist.* 18, 364–371. doi: 10.1089/mdr.2012.0008
- Angen, O., Ahrens, P., and Bisgaard, M. (2002). Phenotypic and genotypic characterization of *Mannheimia* (*Pasteurella*) *haemolytica*-like strains isolated from diseased animals in Denmark. *Vet. Microbiol.* 84, 103–114. doi: 10.1016/S0378-1135(01)00439-4
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Diaz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75



- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Baquero, F. (2011). The 2010 Garrod Lecture: the dimensions of evolution in antibiotic resistance: ex unibus plurum et ex pluribus unum. *J. Antimicrob. Chemother.* 66, 1659–1672. doi: 10.1093/jac/dkr214
- Blanco, M., Gutiérrez-Martín, C. B., Rodríguez-Ferri, E. F., Roberts, M. C., and Navas, J. (2006). Distribution of tetracycline resistance genes in *Actinobacillus pleuropneumoniae* isolates from Spain. *Antimicrob. Agents Chemother.* 50, 702–708. doi: 10.1128/AAC.50.2.702-708.2006
- Blanco, M., Kadlec, K., Gutiérrez-Martín, C. B., de la Fuente, A. J., Schwarz, S., and Navas, J. (2007). Nucleotide sequence and transfer properties of two novel types of *Actinobacillus pleuropneumoniae* plasmids carrying the tetracycline resistance gene *tet(H)*. *J. Antimicrob. Chemother.* 60, 864–867. doi: 10.1093/jac/dkm293
- Bleicher, A., Schoffl, G., Rodicio Mdel, R., and Saluz, H. P. (2013). The plasmidome of a *Salmonella enterica* serovar Derby isolated from pork meat. *Plasmid* 69, 202–210. doi: 10.1016/j.plasmid.2013.01.001
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Borowiak, M., Fischer, J., Hammerl, J. A., Hendriksen, R. S., Szabo, I., and Malorny, B. (2017). Identification of a novel transposon-associated phosphoethanolamine transferase gene, *mcr-5* conferring resistance in d-tartrate fermenting *Salmonella enterica* subsp. *enterica* serovar Paratyphi B. *J. Antimicrob. Chemother.* 72, 3317–3324. doi: 10.1093/jac/dkx327
- Burian, J., Guller, L., Macor, M., and Kay, W. W. (1997). Small cryptic plasmids of multiplasmid, clinical *Escherichia coli*. *Plasmid* 37, 2–14. doi: 10.1006/plas.1996.1273
- Burrough, E. R., Arruda, B. L., and Plummer, P. J. (2017). Comparison of the luminal and mucosa-associated microbiota in the colon of pigs with and without swine dysentery. *Front. Vet. Sci.* 4:139. doi: 10.3389/fvets.2017.00139
- Carattoli, A. (2013). Plasmids and the spread of resistance. *Int. J. Med. Microbiol.* 303, 298–304. doi: 10.1016/j.ijmm.2013.02.001
- Chen, C. Y., Lindsey, R. L., Strobaugh, T. P. Jr., Frye, J. G., and Meinersmann, R. J. (2010). Prevalence of ColE1-like plasmids and kanamycin resistance genes in *Salmonella enterica* serovars. *Appl. Environ. Microbiol.* 76, 6707–6714. doi: 10.1128/AEM.00692-10
- Clinical and Laboratory Standards Institute (CLSI). (2013a). *Performance Standards for Antimicrobial Disk and Dilution Susceptibility Tests for Bacteria Isolated from Animals, Approved Standards—4th Edn*, VET01-A4. Wayne, PA: CLSI.
- Clinical and Laboratory Standards Institute (CLSI). (2013b). *Performance Standards for Antimicrobial Susceptibility Testing; 23rd Edn, Approved Standard M100-S23*. Wayne, PA: CLSI.
- de Toro, M., Rodríguez, I., Rojo-Bezares, B., Helmuth, R., Torres, C., Guerra, B., et al. (2013). pMdT1, a small ColE1-like plasmid mobilizing a new variant of the *aac(6)-Ib-cr* gene in *Salmonella enterica* serovar Typhimurium. *J. Antimicrob. Chemother.* 68, 1277–1280. doi: 10.1093/jac/dkt001
- Durso, L. M., Miller, D. N., and Wienhold, B. J. (2012). Distribution and quantification of antibiotic resistant genes and bacteria across agricultural and non-agricultural metagenomes. *PLoS ONE* 7:e48325. doi: 10.1371/journal.pone.0048325
- Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., and Salmond, G. P. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.* 106, 894–899. doi: 10.1073/pnas.0808832106
- Fraser, J., Donachie, W., Quirie, M., and Gilmour, N. J. (1983). Rapid indirect hemagglutination test for serotyping *Pasteurella haemolytica*. *J. Clin. Microbiol.* 18, 206–207.
- Garbari, L., Busetti, M., Dolzani, L., Petix, V., Knezevich, A., Bressan, R., et al. (2015). pKbuS13, a KPC-2-encoding plasmid from *Klebsiella pneumoniae* sequence type 833, carrying Tn4401b inserted into an Xer site-specific recombination locus. *Antimicrob. Agents Chemother.* 59, 5226–5231. doi: 10.1128/AAC.04543-14
- García-Fernández, A., Fortini, D., Veldman, K., Mevius, D., and Carattoli, A. (2009). Characterization of plasmids harbouring *qnrS1*, *qnrB2* and *qnrB19* genes in *Salmonella*. *J. Antimicrob. Chemother.* 63, 274–281. doi: 10.1093/jac/dkn470
- Garcillán-Barcia, M. P., Francia, M. V., and de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* 33, 657–687. doi: 10.1111/j.1574-6976.2009.00168.x
- Garcillán-Barcia, M. P., Ruiz del Castillo, B., Alvarado, A., de la Cruz, F., and Martínez-Martínez, L. (2015). Degenerate primer MOB typing of multiresistant clinical isolates of *E. coli* uncovers new plasmid backbones. *Plasmid* 77, 17–27. doi: 10.1016/j.plasmid.2014.11.003
- Gregorova, D., Pravcova, M., Karpiskova, R., and Rychlik, I. (2002). Plasmid pC present in *Salmonella enterica* serovar Enteritidis PT14b strains encodes a restriction modification system. *FEMS Microbiol. Lett.* 214, 195–198. doi: 10.1111/j.1574-6968.2002.tb11346.x
- Gupta, V. K., Paul, S., and Dutta, C. (2017). Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front. Microbiol.* 8:1162. doi: 10.3389/fmicb.2017.01162
- Handford, C. L., Stang, C. T., Raivio, T. L., and Dennis, J. J. (2009). The contribution of small cryptic plasmids to the antibiotic resistance of enteropathogenic *Escherichia coli* E2348/69. *Can. J. Microbiol.* 55, 1229–1239. doi: 10.1139/W09-079
- Jacoby, G. A., and Bush, K. (2016). The Curious Case of TEM-116. *Antimicrob. Agents Chemother.* 60:7000. doi: 10.1128/AAC.01777-16
- Koncan, R., Valverde, A., Morosini, M. I., García-Castillo, M., Canton, R., Cornaglia, G., et al. (2007). Learning from mistakes: Taq polymerase contaminated with beta-lactamase sequences results in false emergence of *Streptococcus pneumoniae* containing TEM. *J. Antimicrob. Chemother.* 60, 702–703. doi: 10.1093/jac/dkm239
- Król, J., Bania, J., Florek, M., Pliszczak-Król, A., and Staroniewicz, Z. (2011). Polymerase chain reaction-based identification of clinically relevant Pasteurellaceae isolated from cats and dogs in Poland. *J. Vet. Diagn. Invest.* 23, 532–537. doi: 10.1177/1040638711403434
- Lahlou, H., Dahmen, S., Moussa, M. B., and Omrane, B. (2011). First detection of TEM-116 extended-spectrum beta-lactamase in a *Providencia stuartii* isolate from a Tunisian hospital. *Indian J. Med. Microbiol.* 29, 258–261. doi: 10.4103/0255-0857.83909
- Lancashire, J. F., Terry, T. D., Blackall, P. J., and Jennings, M. P. (2005). Plasmid-encoded Tet B tetracycline resistance in *Haemophilus parasuis*. *Antimicrob. Agents Chemother.* 49, 1927–1931. doi: 10.1128/AAC.49.5.1927-1931.2005
- Lanza, V. F., Baquero, F., Martínez, J. L., Ramos-Ruiz, R., González-Zorn, B., Andremon, A., et al. (2018). In-depth resistome analysis by targeted metagenomics. *Microbiome* 6:11. doi: 10.1186/s40168-017-0387-y
- Lilly, J., and Camps, M. (2015). Mechanisms of theta plasmid replication. *Microbiol. Spectr.* 3:PLAS-0029-2014. doi: 10.1128/microbiolspec.PLAS-0029-2014
- Miranda, C. D., Kehrenberg, C., Ulep, C., Schwarz, S., and Roberts, M. C. (2003). Diversity of tetracycline resistance genes in bacteria from Chilean salmon farms. *Antimicrob. Agents Chemother.* 47, 883–888. doi: 10.1128/AAC.47.3.883-888.2003
- Moleres, J., Santos-Lopez, A., Lazaro, I., Labairu, J., Prat, C., Ardanuy, C., et al. (2015). Novel *blaROB-1*-bearing plasmid conferring resistance to beta-lactams in *Haemophilus parasuis* isolates from healthy weaning pigs. *Appl. Environ. Microbiol.* 81, 3255–3267. doi: 10.1128/AEM.03865-14
- Moran, R. A., and Hall, R. M. (2017). Analysis of pCERC7, a small antibiotic resistance plasmid from a commensal ST131 *Escherichia coli*, defines a diverse group of plasmids that include various segments adjacent to a multimer resolution site and encode the same NikA relaxase accessory protein enabling mobilisation. *Plasmid* 89, 42–48. doi: 10.1016/j.plasmid.2016.11.001
- Pal, C., Bengtsson-Palme, J., Kristiansson, E., and Joakim Larsson, D. G. (2016). The structure and diversity of human, animal and environmental resistomes. *Microbiome* 4, 54. doi: 10.1186/s40168-016-0199-5
- Pallecchi, L., Riccobono, E., Mantella, A., Fernandez, C., Bartalesi, F., Rodriguez, H., et al. (2011). Small *qnrB*-harbouring ColE-like plasmids widespread in commensal enterobacteria from a remote Amazonas population not exposed to antibiotics. *J. Antimicrob. Chemother.* 66, 1176–1178. doi: 10.1093/jac/dkr026
- Pallecchi, L., Riccobono, E., Sennati, S., Mantella, A., Bartalesi, F., Trifoso, C., et al. (2010). Characterization of small ColE-like plasmids mediating widespread

- dissemination of the *qnrB19* gene in commensal enterobacteria. *Antimicrob. Agents Chemother.* 54, 678–682. doi: 10.1128/AAC.01160-09
- Pan, L., Leung, P. C., and Gu, J. D. (2010). A new ColE1-like plasmid group revealed by comparative analysis of the replication proficient fragments of Vibrionaceae plasmids. *J. Microbiol. Biotechnol.* 20, 1163–1178. doi: 10.4014/jmb.1003.03007
- Papagiannitsis, C. C., Dolejska, M., Izdebski, R., Dobiasova, H., Studentova, V., Esteves, F. J., et al. (2015). Characterization of pKP-M1144, a novel ColE1-Like plasmid encoding IMP-8, GES-5, and BEL-1 beta-lactamases, from a *Klebsiella pneumoniae* sequence type 252 isolate. *Antimicrob. Agents Chemother.* 59, 5065–5068. doi: 10.1128/AAC.00937-15
- Roto, S. M., Rubinelli, P. M., and Ricke, S. C. (2015). An introduction to the avian gut microbiota and the effects of yeast-based prebiotic-type compounds as potential feed additives. *Front. Vet. Sci.* 2:28. doi: 10.3389/fvets.2015.00028
- Rozhon, W. M., Petutsching, E. K., and Jonak, C. (2006). Isolation and characterization of pHW15, a small cryptic plasmid from *Rahnella genomospecies* 2. *Plasmid* 56, 202–215. doi: 10.1016/j.plasmid.2006.05.007
- San Millan, A., Escudero, J. A., Catalan, A., Nieto, S., Farelo, F., Gibert, M., et al. (2007). Beta-lactam resistance in *Haemophilus parasuis* is mediated by plasmid pB1000 bearing bla<sub>ROB-1</sub>. *Antimicrob. Agents Chemother.* 51, 2260–2264. doi: 10.1128/AAC.00242-07
- San Millan, A., Escudero, J. A., Guitierrez, B., Hidalgo, L., Garcia, N., Llagostera, M., et al. (2009). Multiresistance in *Pasteurella multocida* is mediated by coexistence of small plasmids. *Antimicrob. Agents Chemother.* 53, 3399–3404. doi: 10.1128/AAC.01522-08
- San Millan, A., Garcia-Cobos, S., Escudero, J. A., Hidalgo, L., Gutierrez, B., Carrilero, L., et al. (2010). *Haemophilus influenzae* clinical isolates with plasmid pB1000 bearing bla<sub>ROB-1</sub>: fitness cost and interspecies dissemination. *Antimicrob. Agents Chemother.* 54, 1506–1511. doi: 10.1128/AAC.01489-09
- San Millan, A., Giufre, M., Escudero, J. A., Hidalgo, L., Gutierrez, B., Cerquetti, M., et al. (2011). Contribution of ROB-1 and PBP3 mutations to the resistance phenotype of a beta-lactamase-positive amoxicillin/clavulanic acid-resistant *Haemophilus influenzae* carrying plasmid pB1000 in Italy. *J. Antimicrob. Chemother.* 66, 96–99. doi: 10.1093/jac/dkq392
- Santos-Lopez, A., Bernabe-Balas, C., Ares-Arroyo, M., Ortega-Huedo, R., Hoefler, A., San Millan, A., et al. (2017). A Naturally occurring single nucleotide polymorphism in a multicopy plasmid produces a reversible increase in antibiotic resistance. *Antimicrob. Agents Chemother.* 61:e01735-16. doi: 10.1128/AAC.01735-16
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Soge, O. O., Adeniyi, B. A., and Roberts, M. C. (2006). New antibiotic resistance genes associated with CTX-M plasmids from uropathogenic *Nigerian Klebsiella pneumoniae*. *J. Antimicrob. Chemother.* 58, 1048–1053. doi: 10.1093/jac/dkl370
- Stoesser, N., Sheppard, A. E., Pankhurst, L., De Maio, N., Moore, C. E., Sebra, R., et al. (2016). Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *MBio* 7:e02162-15. doi: 10.1128/mBio.02162-15
- Stoesser, N., Sheppard, A. E., Peirano, G., Anson, L. W., Pankhurst, L., Sebra, R., et al. (2017). Genomic epidemiology of global *Klebsiella pneumoniae* carbapenemase 334 (KPC)-producing *Escherichia coli*. *Sci. Rep.* 7:5917. doi: 10.1038/s41598-017-06256-2
- Tomizawa, J. (1984). Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript. *Cell* 38, 861–870. doi: 10.1016/0092-8674(84)90281-2
- Tomizawa, J. I., Ohmori, H., and Bird, R. E. (1977). Origin of replication of colicin E1 plasmid DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74, 1865–1869. doi: 10.1073/pnas.74.5.1865
- Townsend, K. M., Boyce, J. D., Chung, J. Y., Frost, A. J., and Adler, B. (2001). Genetic organization of *Pasteurella multocida* cap Loci and development of a multiplex capsular PCR typing system. *J. Clin. Microbiol.* 39, 924–929. doi: 10.1128/JCM.39.3.924-929.2001
- Townsend, K. M., Frost, A. J., Lee, C. W., Papadimitriou, J. M., and Hawkins, H. J. (1998). Development of PCR assays for species- and type-specific identification of *Pasteurella multocida* isolates. *J. Clin. Microbiol.* 36, 1096–1100.
- Tristram, S. G., Littlejohn, R., and Bradbury, R. S. (2010). bla<sub>ROB-1</sub> presence on pB1000 in *Haemophilus influenzae* is widespread, and variable cefaclor resistance is associated with altered penicillin-binding proteins. *Antimicrob. Agents Chemother.* 54, 4945–4947. doi: 10.1128/aac.00263-10
- Vincent, A. T., Emond-Rheault, J. G., Barbeau, X., Attere, S. A., Frenette, M., Lague, P., et al. (2016). Antibiotic resistance due to an unusual ColE1-type replicon plasmid in *Aeromonas salmonicida*. *Microbiology* 162, 942–953. doi: 10.1099/mic.0.000286
- Warburton, P. J., Amodeo, N., and Roberts, A. P. (2016). Mosaic tetracycline resistance genes encoding ribosomal protection proteins. *J. Antimicrob. Chemother.* 71, 3333–3339. doi: 10.1093/jac/dkw304
- Wiedenbeck, J., and Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* 35, 957–976. doi: 10.1111/j.1574-6976.2011.02922.x
- Ye, C., Lan, R., Xia, S., Zhang, J., Sun, Q., Zhang, S., et al. (2010). Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of *Shigella flexneri*. *J. Clin. Microbiol.* 48, 419–426. doi: 10.1128/JCM.00614-09
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261
- Zeil, C., Widmann, M., Fadernrecht, S., Vogel, C., and Pleiss, J. (2016). Network analysis of sequence-function relationships and exploration of sequence space of TEM beta-lactamases. *Antimicrob. Agents Chemother.* 60, 2709–2717. doi: 10.1128/AAC.02930-15

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ares-Arroyo, Bernabe-Balas, Santos-Lopez, Baquero, Prasad, Cid, Martin-Espada, San Millan and Gonzalez-Zorn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal *Salmonella enterica*

Saskia Neuert<sup>1,2</sup>, Satheesh Nair<sup>2</sup>, Martin R. Day<sup>2</sup>, Michel Doumith<sup>2</sup>, Philip M. Ashton<sup>2</sup>, Kate C. Mellor<sup>3,4</sup>, Claire Jenkins<sup>1,2</sup>, Katie L. Hopkins<sup>2</sup>, Neil Woodford<sup>2</sup>, Elizabeth de Pinna<sup>2</sup>, Gauri Godbole<sup>1,2</sup> and Timothy J. Dallman<sup>1,2\*</sup>

<sup>1</sup> National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, Liverpool, United Kingdom, <sup>2</sup> Bacteriology Reference Department, National Infection Service, Public Health England, London, United Kingdom, <sup>3</sup> Department of Pathobiology and Population Sciences, Royal Veterinary College, London, United Kingdom, <sup>4</sup> London School of Hygiene & Tropical Medicine, London, United Kingdom

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Sultan Zainal Abidin University,  
Malaysia

### Reviewed by:

Jeanette Teo,  
National University Hospital,  
Singapore  
Debashree Basu,  
University of Minnesota Twin Cities,  
United States

### \*Correspondence:

Timothy J. Dallman  
tim.dallman@phe.gov.uk

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 18 December 2017

**Accepted:** 15 March 2018

**Published:** 27 March 2018

### Citation:

Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, Jenkins C, Hopkins KL, Woodford N, de Pinna E, Godbole G and Dallman TJ (2018) Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal *Salmonella enterica*. *Front. Microbiol.* 9:592. doi: 10.3389/fmicb.2018.00592

Surveillance of antimicrobial resistance (AMR) in non-typhoidal *Salmonella enterica* (NTS), is essential for monitoring transmission of resistance from the food chain to humans, and for establishing effective treatment protocols. We evaluated the prediction of phenotypic resistance in NTS from genotypic profiles derived from whole genome sequencing (WGS). Genes and chromosomal mutations responsible for phenotypic resistance were sought in WGS data from 3,491 NTS isolates received by Public Health England's Gastrointestinal Bacteria Reference Unit between April 2014 and March 2015. Inferred genotypic AMR profiles were compared with phenotypic susceptibilities determined for fifteen antimicrobials using EUCAST guidelines. Discrepancies between phenotypic and genotypic profiles for one or more antimicrobials were detected for 76 isolates (2.18%) although only 88/52,365 (0.17%) isolate/antimicrobial combinations were discordant. Of the discrepant results, the largest number were associated with streptomycin (67.05%,  $n = 59$ ). Pan-susceptibility was observed in 2,190 isolates (62.73%). Overall, resistance to tetracyclines was most common (26.27% of isolates,  $n = 917$ ) followed by sulphonamides (23.72%,  $n = 828$ ) and ampicillin (21.43%,  $n = 748$ ). Multidrug resistance (MDR), i.e., resistance to three or more antimicrobial classes, was detected in 848 isolates (24.29%) with resistance to ampicillin, streptomycin, sulphonamides and tetracyclines being the most common MDR profile ( $n = 231$ ; 27.24%). For isolates with this profile, all but one were *S. Typhimurium* and 94.81% ( $n = 219$ ) had the resistance determinants *bla*<sub>TEM-1</sub>, *strA-strB*, *sul2* and *tet(A)*. Extended-spectrum  $\beta$ -lactamase genes were identified in 41 isolates (1.17%) and multiple mutations in chromosomal genes associated with ciprofloxacin resistance in 82 isolates (2.35%). This study showed that WGS is suitable as a rapid means of determining AMR patterns of NTS for public health surveillance.

**Keywords:** antimicrobial resistance, multidrug resistance, non-typhoidal *Salmonella enterica*, whole genome sequencing, public health surveillance, One Health

## INTRODUCTION

*Salmonella enterica* subspecies *enterica* is responsible for 99% of salmonellosis cases in humans and animals, and can be further subdivided into the host-restricted typhoidal salmonellae and the more generalist non-typhoidal salmonellae (NTS) (Langridge et al., 2015; Wain et al., 2015). As host-adapted or generalist organisms, NTS can be transferred from animals to humans causing zoonotic infections and therefore fall under the World Health Organization's One Health approach. Globally, NTS were estimated to cause 93.8 million enteric infections resulting in 155,000 deaths annually (Majowicz et al., 2010), and in the United Kingdom they are the third most common cause of bacterial gastroenteritis (Tam et al., 2012). Although NTS symptoms are often limited to the gastrointestinal tract, invasive disease can occur, especially in high-risk groups such as immunocompromised patients and the elderly (Parry et al., 2013). Invasive disease has also been observed in several low-income settings (Kingsley et al., 2009; Feasey et al., 2016; Ashton et al., 2017) and was estimated to result in 3.4 million cases and 681,000 deaths worldwide in 2010, with the heaviest burden on the African continent (Ao et al., 2015).

While the use of antimicrobial agents to treat invasive and severe gastrointestinal cases has decreased mortality rates for NTS infections, and veterinary antimicrobial therapy has lowered the risk of zoonoses, these interventions have come with a price. Increased use of the traditional first-line drugs ampicillin, chloramphenicol, streptomycin, sulphonamides and tetracycline quickly led to the emergence of ACSSuT-type *S. enterica* serovar Typhimurium strains in the 1980s, resistant to exactly these drugs (Threlfall et al., 1996; Boyd et al., 2002). Resistance to fluoroquinolones, introduced to circumvent this problem, developed as a consequence of the veterinary use of enrofloxacin (Threlfall et al., 1997). NTS strains resistant to extended-spectrum cephalosporins, an alternative to fluoroquinolones for the treatment of invasive disease, have been detected throughout Europe since the 1990s (Tassios et al., 1999; Villa et al., 2002; Burke et al., 2014). By 2015, 29.3% of the NTS isolates in the European Union were categorized as multidrug-resistant (MDR) (EFSA, 2017). More recently, the spread of an extensively drug-resistant strain of *S. Kentucky*, non-susceptible to ciprofloxacin, extended-spectrum cephalosporins, carbapenems, most aminoglycosides, trimethoprim-sulfamethoxazole, and azithromycin, has sparked concern (Le Hello et al., 2013). Resistance to azithromycin has been reported in other NTS serovars (Villa et al., 2015; Nair et al., 2016). Acquired resistance to colistin, considered the antimicrobial of last resort for the treatment of many MDR Gram-negative pathogens, has also been detected in NTS (Doumith et al., 2016).

Due to the association of MDR NTS infection with increased mortality and higher costs to the healthcare system, determination of antimicrobial resistance (AMR) profiles is an essential part of NTS surveillance in reference laboratories. Phenotypic serotyping and phage typing at Public Health England's (PHE) Gastrointestinal Bacteria Reference Unit (GBRU) has been replaced by the routine implementation

of whole genome sequencing (WGS) for identification and surveillance of *Salmonella* since April 2014 (Ashton et al., 2016). As well as providing information about phylogenetic relationships between isolates, the sequencing data can be used to identify resistance determinants and therefore constitutes a rapid alternative to monitor emerging trends in AMR patterns of NTS. With this study, we sought to evaluate the suitability of inferring AMR profiles from genotype in NTS in comparison with traditional phenotypic susceptibility testing.

## MATERIALS AND METHODS

### Bacterial Isolates

Between April 2014 and March 2015, PHE received 7,009 NTS *S. enterica* subspecies *enterica* isolates for surveillance purposes. After deduplication of outbreak cases and exclusion of isolates with WGS of insufficient quality, results of phenotypic susceptibility testing and genotypic profiling were available for 3,491 isolates (49.81%). These comprised 227 different serovars plus 66 isolates that could not successfully be subtyped to serovar level. GBRU's routine phenotypic testing strategy for surveillance of NTS attempts to maximize the detection of AMR by focussing on serovars known to have high resistance rates. This leads to an under-representation of some serovars, such as *S. Enteritidis*, and an over-representation of others, such as *S. Infantis* and *S. Kentucky*, in this dataset. Amongst the isolates included in the analysis, the ten most common serovars were *S. Typhimurium* (23.69%,  $n = 827$ ), *S. Enteritidis* (8.42%,  $n = 294$ ), *S. Virchow* (4.01%,  $n = 140$ ), *S. Stanley* (3.98%,  $n = 139$ ), *S. Newport* (3.75%,  $n = 131$ ), *S. Infantis* (3.47%,  $n = 121$ ), *S. Kentucky* (3.12%,  $n = 109$ ), *S. Oranienburg* (2.06%,  $n = 72$ ), *S. Java* (2.03%,  $n = 71$ ) and *S. Saint-Paul* (1.78%,  $n = 62$ ). The majority ( $n = 3487$ ) of isolates were of human origin, three were derived from food and one from an unknown source.

### Whole Genome Sequencing

Sequencing libraries were prepared from extracted genomic DNA using the Nextera XT DNA Sample Preparation kit (Illumina, Cambridge, United Kingdom). Short-read sequence fragments of 100 bp were produced by paired-end sequencing on an Illumina HiSeq platform (Illumina, Cambridge, United Kingdom). FASTQ sequences were deposited in the NCBI Short Read Archive under the BioProject PRJNA315192. Short read archive accession numbers are available in **Supplementary Table S1**.

### Serovar Prediction

Serovars were inferred from the sequencing data using the seven-gene MLST and eBurst Group approach (Achtman et al., 2012; Ashton et al., 2016). Traditional serotyping was not performed.

### Detection of Antimicrobial Resistance Determinants

For the identification of AMR determinants, the 'Genefinder' algorithm was employed, which maps the sequencing reads to a set of reference sequences using Bowtie 2 followed by generation



of an mpileup file using Samtools (Langmead and Salzberg, 2012). To establish the presence of the reference sequence or nucleotide variations within the read set, a positive match had to meet the following criteria: query coverage 100%, base-call variation > 85% and nucleotide identity > 90%.

The reference database used included acquired genes and mutations known to confer resistance to  $\beta$ -lactams (including penicillins, 2nd-, 3rd- and 4th-generation cephalosporins and carbapenems), fluoroquinolones, aminoglycosides, sulphonamides, tetracyclines, trimethoprim and phenicols (Day et al., 2017b; Sadouki et al., 2017). Variants of  $\beta$ -lactamase genes were identified with 100% identity based on reference sequences downloaded from the Lahey<sup>1</sup> or NCBI  $\beta$ -lactamase data resources<sup>2</sup>. Further reference sequences for acquired resistance genes were obtained from the Comprehensive Antimicrobial Resistance Database<sup>3</sup> and the Resfinder datasets<sup>4</sup>. Chromosomal mutations were limited to previously published variations within the quinolone resistance-determining regions (QRDRs) of *gyrA* and *parC*.

## Antimicrobial Susceptibility Testing

Isolates were recovered from the PHE archive and retrospective susceptibility testing was performed and interpreted using EUCAST breakpoints and screening concentrations<sup>5</sup>. For the purpose of epidemiologically screening the large numbers of *S. enterica* isolates received by the reference laboratory, agar dilution with Mueller–Hinton agar was used to determine breakpoint values of ampicillin, cefotaxime, ceftazidime, cefpirome, ertapenem, chloramphenicol, gentamicin, streptomycin, tobramycin, sulphonamides, tetracycline, trimethoprim and ciprofloxacin. Decreased susceptibility (MIC 0.06–0.25 mg/L) and resistance (MIC > 0.5 mg/L) were distinguished for ciprofloxacin. If required, MICs were confirmed by Etest® (bioMérieux, Marcy-l'Étoile, France) or by agar dilution. To aid detection of OXA-48-like carbapenemases and acquired AmpC genes, breakpoint testing of temocillin and ceftoxitin, respectively, was included in the panel.

## Statistical Analysis

Comparisons were made between the prevalence of resistance determinants associated with isolates, for which a travel history was available, and those for which there was no information about recent travel. Travel destinations were grouped according to the United Nations geoscheme. Statistical significance was assessed using the chi-square test. A  $p$ -value  $\leq 0.05$  was considered statistically significant. Statistical analysis was performed using R's `chisq.test` function.

<sup>1</sup> www.lahey.org

<sup>2</sup> https://www.ncbi.nlm.nih.gov/pathogens/beta-lactamase-data-resources

<sup>3</sup> http://arpcard.mcmaster.ca

<sup>4</sup> https://cge.cbs.dtu.dk/services/data.php

<sup>5</sup> http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST\_files/Breakpoint\_tables/v\_8.0\_Breakpoint\_Tables.pdf

## RESULTS

### Comparison Between Phenotypic and Genotypic AMR Profiles

Phenotypic and genotypic AMR profiling was highly correlated, with the profiles of 3,415 isolates (97.82%) being entirely in agreement for both approaches for all 15 antimicrobials from nine different classes. For the 76 isolates with discordant results, the genotype wrongly predicted pan-susceptibility for one isolate (1.32%). This isolate was phenotypically resistant to one antimicrobial. For a further 64 discrepant isolates (84.21%), the mismatch was based on false or missing prediction of resistance to a single antimicrobial, and for 11 (14.47%) based on two antimicrobials.

Overall, 88 (0.17%) out of a possible 52,365 isolate/antimicrobial combinations did not match (Table 1). Of these discrepant results, 69/88 (78.41%) constituted major errors (MEs), i.e., isolates were genotypically predicted to be resistant but showed phenotypic susceptibility, rather than very major errors (VMEs), which were genotypically susceptible but phenotypically resistant. The largest fraction of the 88 mismatches could be attributed to streptomycin ( $n = 59$ , 67.05%), 51 of these were MEs. Sensitivity of resistance prediction from genotype was  $\geq 95\%$  for all antimicrobials except temocillin. However, only a single isolate was found to be phenotypically resistant to temocillin. Specificity of prediction exceeded 98% for all fifteen antimicrobials tested.

### Resistance to $\beta$ -lactams

Of the 3,491 isolates in this study, 749 (21.46%) carried genes conferring resistance to  $\beta$ -lactam antibiotics (Supplementary Table S2). The most common genes were the penicillinase-encoding *bla*<sub>TEM-1</sub> ( $n = 603$ ) and *bla*<sub>PSE-1</sub>/*bla*<sub>CARB-2</sub> ( $n = 75$ ). Additionally, other TEM-type  $\beta$ -lactamase genes were detected in 36 isolates, including *bla*<sub>TEM-117</sub> ( $n = 12$ ) and *bla*<sub>TEM-135</sub> ( $n = 7$ ). The single ME associated with predicted ampicillin resistance was due to the presence of *bla*<sub>TEM-1</sub> without phenotypic consequences. Seven isolates (0.20%) carried OXA-type class D  $\beta$ -lactamases. Of these, four were found in *S. Typhimurium* and two in *S. Kentucky*.

Genes for CTX-M-type extended-spectrum  $\beta$ -lactamases (ESBLs) were present in 41 isolates (1.17%), most commonly *bla*<sub>CTX-M-9</sub> ( $n = 10$ ) and *bla*<sub>CTX-M-55</sub> ( $n = 9$ ). Twenty of these were *S. Typhimurium* and five *S. Kentucky*. Additionally, four isolates carried the *bla*<sub>SHV-12</sub> ESBL gene. No ESBL genes were detected in *S. Enteritidis*. Combinations of penicillinase and ESBL genes occurred in 16 isolates, most frequently *bla*<sub>TEM-1</sub> with *bla*<sub>CTX-M-55</sub> ( $n = 9$ ).

Sixteen isolates (0.46%), seven of these *S. Typhimurium* and one *S. Kentucky*, had the acquired AmpC resistance gene *bla*<sub>CMY-2</sub>. Carbapenemase genes were not detected.

### Resistance to Quinolones

Multiple mutations in the QRDR of the DNA gyrase subunit gene *gyrA* in combination with multiple mutations in the DNA topoisomerase gene *parC* are expected to confer full ciprofloxacin

**TABLE 1** | Comparison of phenotypic antimicrobial susceptibility testing and genome-derived resistance prediction for non-typhoidal *Salmonella enterica* ( $n = 3491$ ).

Antimicrobial	Phenotype: susceptible		Phenotype: resistant		Sensitivity (%)	Specificity (%)
	Genotype: resistant	Genotype: susceptible	Genotype: resistant	Genotype: susceptible		
Ampicillin	1	2742	747	1	99.87	99.96
Temocillin	0	3490	0	1	0	100
Cefoxitin	0	3471	19	1	95.0	100
Cefotaxime	0	3434	57	0	100	100
Ceftazidime	0	3444	47	0	100	100
Cefpirome	0	3444	47	0	100	100
Ertapenem	0	3481	10	0	100	100
Chloramphenicol	4	3284	201	2	99.01	99.88
Gentamicin	1	3351	138	1	99.28	99.97
Streptomycin	51	2821	613	8	98.71	98.22
Tobramycin	2	3392	97	0	100	99.94
Sulphonamides	2	2661	828	0	100	99.92
Tetracycline	6	2568	917	0	100	99.77
Trimethoprim	1	3185	301	4	98.69	99.97
Ciprofloxacin	1	3352	137	1	99.28	99.97

Values shown designate the number of isolates. For ciprofloxacin only isolates with a MIC > 0.5 mg/L are shown.

resistance (MIC > 0.5 mg/L) and were observed in 82 isolates (2.35%) (Table 2). The most common combinations were either *gyrA*[83:S-F;87:D-Y] ( $n = 41$ ) or *gyrA*[83:S-F;87:D-N] ( $n = 25$ ) in conjunction with *parC*[57:T-S;80:S-I]. For *S. Kentucky*, multiple QRDR mutations were identified in 77 isolates. Neither *S. Typhimurium* nor *S. Enteritidis* had any of these mutations (Supplementary Table S2).

A further 599 isolates (17.16%) harbored determinants responsible for reduced susceptibility to ciprofloxacin (MIC 0.06–0.25 mg/L) with or without *parC* mutations. These included a single *gyrA* mutation in the QRDR ( $n = 430$ ), most commonly *gyrA*[87:D-Y] ( $n = 155$ ) or *gyrA*[83:S-Y] ( $n = 112$ ) and/or one or multiple plasmid-mediated quinolone resistance (PMQR) genes ( $n = 195$ ). The most frequent PMQR genes detected were *qnrS1* ( $n = 95$ ) and *qnrB19* ( $n = 49$ ). PMQR genes were rare in *S. Kentucky* ( $n = 1$  compared with  $n = 188$  for chromosomal mutations). One or more PMQR determinants in combination with a single *gyrA* mutation were found in twenty isolates. Of the isolates carrying both multiple *parC* and *gyrA* mutations, only one *S. Indiana* had additional PMQR genes, namely the efflux pump-encoding *oqxA* and *oqxB*.

Seven isolates carried the fluoroquinolone- and aminoglycoside-modifying *N*-acetyltransferase gene variant *aac(6')-Ib-cr*, six of these in combination with other quinolone resistance determinants. Of the 138 isolates showing full ciprofloxacin resistance, nineteen carried a single *gyrA* mutation only, 17 a single *gyrA* mutation together with a PMQR gene, 20 had one or more PMQR genes and a single isolate carried *parC*[57:T-S] only (Table 2). The single ME associated with predicted ciprofloxacin resistance was based on the presence of *gyrA*[83:S-F;87:D-N] and *parC*[57:T-S;80:S-I] resulting in reduced susceptibility instead of full resistance.

## Resistance to Aminoglycosides

Genes predicted to confer resistance to streptomycin were detected in 728 isolates (20.85%): 436 had *strA-strB* only and 292 carried genes encoding aminoglycoside adenyltransferases, most commonly *aadA2* ( $n = 189$ ) and *aadA17* ( $n = 107$ ) (Supplementary Table S2). Both *strA-strB* and an *aadA* variant were observed in 101 isolates. Of the 51 MEs associated with streptomycin resistance, 27 were due to the presence of *strA-strB* and twelve had *aadA2* and *aadA17* without phenotypic consequences.

All but eight of the total 3,491 isolates carried an aminoglycoside acetyltransferase *aac(6')*-type gene. However, the majority either had the *aac(6')-Iy* ( $n = 1997$ ), more common in *S. Enteritidis* ( $n = 297$ ), or *aac(6')-Iaa* variant ( $n = 1486$ ), more common in *S. Typhimurium* ( $n = 869$ ) and *S. Kentucky* ( $n = 81$ ). Of the 2,726 isolates carrying either of these two genes as the only aminoglycoside resistance determinant, only eleven showed phenotypic resistance to an aminoglycoside antimicrobial.

Aminoglycoside acetyltransferase *aac(3)* variants associated with resistance to gentamicin and tobramycin were detected in 130 isolates (3.72%), most notably *aac(3)-Id* ( $n = 50$ ) and *aac(3)-IIa* ( $n = 36$ ). *aac(3)-IVa*, which confers resistance to the veterinary aminoglycoside apramycin, was present in 24 isolates. No *aac(3)* variants were found in *S. Enteritidis*. Furthermore, the aminoglycoside adenyltransferase gene *ant(2'')-Ia* ( $n = 12$ ) and the aminoglycoside phosphotransferase genes *aph(4)-Ia* ( $n = 23$ ) and *aph(3')-IIa* ( $n = 10$ ) were identified. None of these were present in *S. Enteritidis* or *S. Kentucky*. No 16S rRNA methyltransferase genes were detected. In the single isolate predicted to be resistant to gentamicin but showing phenotypic susceptibility, *aac(3)-IId* was observed. For prediction of tobramycin resistance, one ME was associated with the presence of *ant(2'')-Ia* and the second one with *aac(3)-IIa*.

**TABLE 2 |** Relationship between decreased ciprofloxacin susceptibility (<CIP, MIC 0.06–0.25 mg/L), full ciprofloxacin resistance (>CIP, MIC > 0.5 mg/L) and the most common genotypic quinolone resistance determinants in non-typhoidal *Salmonella enterica*.

Number of isolates	Phenotype				Genotype
	<CIP		>CIP		
	S	R	S	R	
116	6	110	114	2	<i>gyrA</i> [87:D-Y]
51	1	50	42	9	<i>gyrA</i> [83:S-Y]; <i>parC</i> [57:T-S]
51	4	47	51	0	<i>gyrA</i> [87:D-N]
47	0	47	43	4	<i>gyrA</i> [83:S-Y]
44	0	44	39	5	<i>parC</i> [57:T-S]; <i>qnrS1</i>
41	0	41	0	41	<i>gyrA</i> [83:S-F;87:D-Y]; <i>parC</i> [57:T-S;80:S-I]
37	1	36	28	9	<i>qnrS1</i>
34	2	32	34	0	<i>gyrA</i> [87:D-Y]; <i>parC</i> [57:T-S]
32	0	32	31	1	<i>parC</i> [57:T-S]; <i>qnrB19</i>
29	1	28	28	1	<i>gyrA</i> [83:S-F]; <i>parC</i> [57:T-S]
25	0	25	1	24	<i>gyrA</i> [83:S-F;87:D-N]; <i>parC</i> [57:T-S;80:S-I]
23	0	23	20	3	<i>gyrA</i> [83:S-F]
19	2	17	19	0	<i>gyrA</i> [87:D-G]
17	2	15	17	0	<i>gyrA</i> [87:D-G]; <i>parC</i> [57:T-S]
12	6	6	12	0	<i>parC</i> [57:T-S]; <i>qnrD</i>
12	0	12	9	3	<i>qnrB19</i>
11	0	11	0	11	<i>gyrA</i> [83:S-F;87:D-G]; <i>parC</i> [57:T-S;80:S-I]
11	0	11	11	0	<i>gyrA</i> [87:D-N]; <i>parC</i> [57:T-S]
8	0	8	8	0	<i>qnrA1</i>
7	1	6	3	4	<i>aac</i> (6')-Ib-cr
4	0	4	0	4	<i>gyrA</i> [83:S-Y]; <i>parC</i> [57:T-S]; <i>qnrS1</i>
4	0	4	0	4	<i>gyrA</i> [87:D-Y]; <i>qnrS1</i>
4	0	4	4	0	<i>parC</i> [57:T-S]; <i>qnrB9</i>
4	0	4	4	0	<i>parC</i> [57:T-S]; <i>qnrS2</i>
4	0	4	2	2	<i>qnrB6</i>
3	0	3	0	3	<i>gyrA</i> [83:S-F;87:D-G]; <i>parC</i> [57:T-S;80:S-R]
3	0	3	1	2	<i>gyrA</i> [83:S-Y]; <i>parC</i> [57:T-S]; <i>qnrD</i>
3	0	3	0	3	<i>gyrA</i> [83:S-Y]; <i>qnrS1</i>
2	0	2	2	0	<i>parC</i> [57:T-S]; <i>qnrA1</i>
2	0	2	2	0	<i>parC</i> [57:T-S]; <i>qnrB1</i>

Values shown are the number of isolates. S, susceptible; R, Resistant.

## Resistance to Sulphonamides, Tetracyclines and Trimethoprim

Sulphonamide resistance genes were found in 830 isolates (23.78%): 490 carried *sul2*, 350 *sul1* and 75 *sul3* (Supplementary Table S2). Seventy-seven isolates had a combination of two different *sul* genes, most notably *sul1* and *sul2* ( $n = 37$ ), and four isolates carried all three variants. Of the two MEs that occurred for the prediction of sulphonamide resistance, one was based on the presence of *sul2* and one on the presence of *sul1* without phenotypic consequences.

Tetracycline resistance genes occurred in 927 isolates (26.55%), mostly *tet*(A) ( $n = 843$ ). Additional, less frequently encountered genes were the efflux pump-encoding *tet*(G) ( $n = 68$ ), *tet*(C) ( $n = 10$ ) and *tet*(D) ( $n = 5$ ), and the ribosomal protection protein-producing *tet*(M) ( $n = 57$ ). Fifty-six isolates carried a combination of two different genes, mainly *tet*(A) and

*tet*(M) ( $n = 51$ ). Five of the six isolates with predicted but not phenotypic tetracycline resistance harbored *tet*(M).

Trimethoprim resistance-conferring *dfrA* gene variants were identified in 302 isolates (8.65%), most commonly *dfrA12* ( $n = 84$ ), *dfrA1* ( $n = 81$ ) and *dfrA14* ( $n = 65$ ). The remaining isolates carried eight additional variants of *dfrA*. Only one isolate harbored a combination of two different genes (*dfrA1*;*dfrA12*). The single ME associated with prediction of trimethoprim resistance was due to the presence of *dfrA14* without phenotypic consequences.

## Resistance to Phenicol

Genes linked to chloramphenicol resistance were identified in 215 isolates (6.16%) (Supplementary Table S2). Efflux pump genes were found in 194 isolates: *floR* ( $n = 147$ ) and/or *cmlA1* ( $n = 67$ ). All four MEs were associated with the

presence of *cmlA1*. Chloramphenicol acetyltransferase genes of the *catA*- or *catB*-type were detected in 32 isolates. Eleven isolates harbored genes encoding both an efflux pump and an acetyltransferase.

## Multidrug Resistance

Out of a total 3,491 isolates, 1,301 (37.27%) were phenotypically resistant to at least one antimicrobial of the testing panel (Table 3). For the two most common serovars, *S. Typhimurium* and *S. Enteritidis*, this applied to 568/827 (68.68%) and 130/294 (44.22%) isolates, respectively, and for *S. Kentucky* to 82/109 isolates (75.23%).

MDR, i.e., resistance to three or more antimicrobial classes, was observed in 848 of all the NTS isolates (24.29%), 467 *S. Typhimurium* (56.47%), 70 *S. Kentucky* (64.22%) and only 13 *S. Enteritidis* (4.42%). One *S. Typhimurium* isolate exhibited resistance to all nine antimicrobial classes tested.

Detected in 231 isolates overall (6.62%), resistance to ampicillin, streptomycin, sulphonamides and tetracyclines was the most commonly occurring MDR profile, all but one isolate were *S. Typhimurium*. In 219 isolates with this profile, the underlying genotype was *bla*<sub>TEM-1</sub>, *strA-strB sul2*, *tet(A)*. For *S. Enteritidis*, decreased susceptibility to ciprofloxacin was observed most frequently ( $n = 87$ ) with *gyrA*[87:D-Y] being the

**TABLE 3 |** Most common combinations of antimicrobial resistance phenotypes and genotypes in non-typhoidal *Salmonella enterica* for all serovars, *S. Typhimurium* and *S. Enteritidis*.

Serovar	Antimicrobial classes	Number of isolates (%)	Most common phenotypic combination (number of isolates)	Most common genotypic combination (number of isolates)
Total ( $n = 3491$ )	0	2190 (62.73)	–	–
	1/2	453 (12.98)	<CIP (196)	<i>gyrA</i> [87:D-Y] (52)
	3/4	514 (14.72)	AMP/STR/SUL/TET (231)	<i>sul2;strA-strB;tet(A);bla</i> <sub>TEM-1</sub> (219)
	5/6/7	315 (9.02)	AMP/CHL/STR/SUL/TET (37)	<i>sul1;tet(G);aadA17;aadA2;floR;bla</i> <sub>PSE-1/bla</sub> <sub>CARB-2</sub> (30)
	8/9	19 (0.54)	AMP/CAZ/CHL/ < CIP/CPR/CTX/ETP/ FOX/GEN/STR/SUL/TET/ TMP/TOB (1)	<i>qnrS1;sul1;sul2;sul3;strA-strB;tet(M);tet(A);dfrA12;aac(3)-Iva;aadA2;aadA12;aph(4)-Ia;cml1;floR;bla</i> <sub>CMY-2</sub> ;bla <sub>TEM-1</sub> (1)
<i>S. Typhimurium</i> ( $n = 827$ )	0	259 (31.32)	–	–
	1/2	101 (12.21)	TET (40)	<i>tet(A)</i> (38)
	3/4	317 (38.33)	AMP/STR/SUL/TET (230)	<i>sul2;strA-strB;tet(A);bla</i> <sub>TEM-1</sub> (220)
	5/6/7	138 (16.69)	AMP/CHL/STR/SUL/TET (33)	<i>sul1;tet(G);aadA17;aadA2;floR;bla</i> <sub>PSE-1/bla</sub> <sub>CARB-2</sub> (28)
	8/9	12 (1.45)	AMP/CAZ/CHL/ < CIP/CPR/CTX/ETP/ FOX/GEN/STR/SUL/TET/ TMP/TOB (1)	<i>qnrS1;sul1;sul2;sul3;strA-strB;tet(M);tet(A);dfrA12;aac(3)-Iva;aadA2;aadA12;aph(4)-Ia;cml1;floR;bla</i> <sub>CMY-2</sub> ;bla <sub>TEM-1</sub> (1)
<i>S. Enteritidis</i> ( $n = 294$ )	0	164 (55.78)	–	–
	1/2	117 (39.80)	<CIP (87)	<i>gyrA</i> [87:D-Y] (24)
	≥3	13(4.42)	AMP/ < CIP/STR/SUL/TET (4)	<i>gyrA</i> [87:D-N]; <i>sul2;strA-strB;tet(A);bla</i> <sub>TEM-1</sub> (3)
<i>S. Kentucky</i> ( $n = 109$ )	0	27 (24.77)	–	–
	1/2	12 (11.01)	AMP/> CIP (5)	<i>gyrA</i> [83:S-F;87:D-Y]; <i>parC</i> [57:T-S;80:S-I];bla <sub>TEM-1</sub> (4)
	3/4	19 (17.43)	AMP/> CIP/SUL/TET (11)	<i>gyrA</i> [83:S-F;87:D-Y]; <i>parC</i> [57:T-S;80:S-I]; <i>sul1;tet(A);aadA7;bla</i> <sub>TEM-1</sub> (3)
	5/6/7	51 (46.79)	AMP/> CIP/GEN /STR/SUL/TET (24)	<i>gyrA</i> [83:S-F;87:D-Y]; <i>parC</i> [57:T-S;80:S-I]; <i>sul1;tet(A);aac(3)-Id;aadA7;bla</i> <sub>TEM-1</sub> (20)

AMP, ampicillin; FOX, cefoxitin; CTX, cefotaxime; CAZ, ceftazidime; CPR, cefpirome; ETP, ertapenem; CHL, chloramphenicol; GEN, gentamicin; STR, streptomycin; TOB, tobramycin; SUL, sulphonamides; TET, tetracycline; TMP, trimethoprim; <CIP, ciprofloxacin MIC 0.06–0.25 mg/L; >CIP, ciprofloxacin MIC > 0.5 mg/L. Values in the column antimicrobial class denote the numbers of antimicrobial classes the isolates are resistant to.



most common genotypic determinant ( $n = 24$ ). The majority of resistant *S. Kentucky* showed phenotypic resistance to ampicillin, ciprofloxacin, gentamicin, streptomycin, sulphonamides and tetracyclines ( $n = 24$ ). In 20 isolates, this profile was based on the presence of *bla*<sub>TEM-1</sub>, *gyrA*[83:S-F;87:D-Y], *parC*[57:T-S;80:S-I], *aac*(3)-*Id*, *aadA7*, *sul1* and *tet*(A) ( $n = 20$ ). Thirty-three *S. Typhimurium* isolates (3.99%) exhibited the penta-resistant phenotype with resistance to ampicillin, chloramphenicol, streptomycin, sulphonamides and tetracyclines. Of these, 28 carried a combination of *bla*<sub>PSE-1</sub>/*bla*<sub>CARB-2</sub>, *floR*, *aadA17*, *aadA2*, *sul1* and *tet*(G).

## AMR and International Travel

Travel history data was available for 1,070 isolates (30.65%) (Supplementary Table S3). The proportion of isolates resistant to at least one antimicrobial of the testing panel was significantly higher for isolates known to be travel-associated ( $p = 4.5 \times 10^{-5}$ ) (Figure 1). MDR, on the other hand, was correlated with travel to specific regions, namely Eastern Africa ( $p = 0.04$ ), North Africa ( $p = 0.005$ ), Western Africa ( $p = 0.03$ ), Southeast Asia ( $p = 3.4 \times 10^{-5}$ ) and the Caribbean ( $p = 1.3 \times 10^{-4}$ ). ESBL genes were more likely to be found in isolates related to travel to North Africa ( $p = 0.01$ ) and South America ( $p = 0.03$ ). Mutations and acquired genes conferring decreased susceptibility or resistance to ciprofloxacin were more likely to occur in travel-associated isolates (resistance-conferring mutations:  $p = 8.5 \times 10^{-6}$ ; single *gyrA* mutations:  $p = 4.5 \times 10^{-9}$ ; PMQRs:  $p = 1.3 \times 10^{-7}$ ). The presence of genes conferring ciprofloxacin resistance was associated with travel to Southern Asia ( $p = 5.6 \times 10^{-6}$ ). Determinants of aminoglycoside resistance were more prevalent in travel-related isolates, particularly for travel destinations in North Africa, Asia and the Caribbean. The presence of sulphonamide and tetracycline resistance genes was linked to travel to Southeast and Western Asia and the Caribbean while *dfpA* genes were commonly found in isolates associated with travel to North Africa ( $p = 5.2 \times 10^{-7}$ ) and South Asia ( $p = 0.01$ ). Furthermore, travel to North Africa or Southeast Asia was a risk factor for acquisition of isolates carrying chloramphenicol resistance genes ( $p = 0.009$  and  $p = 4.7 \times 10^{-10}$ , respectively).

## DISCUSSION

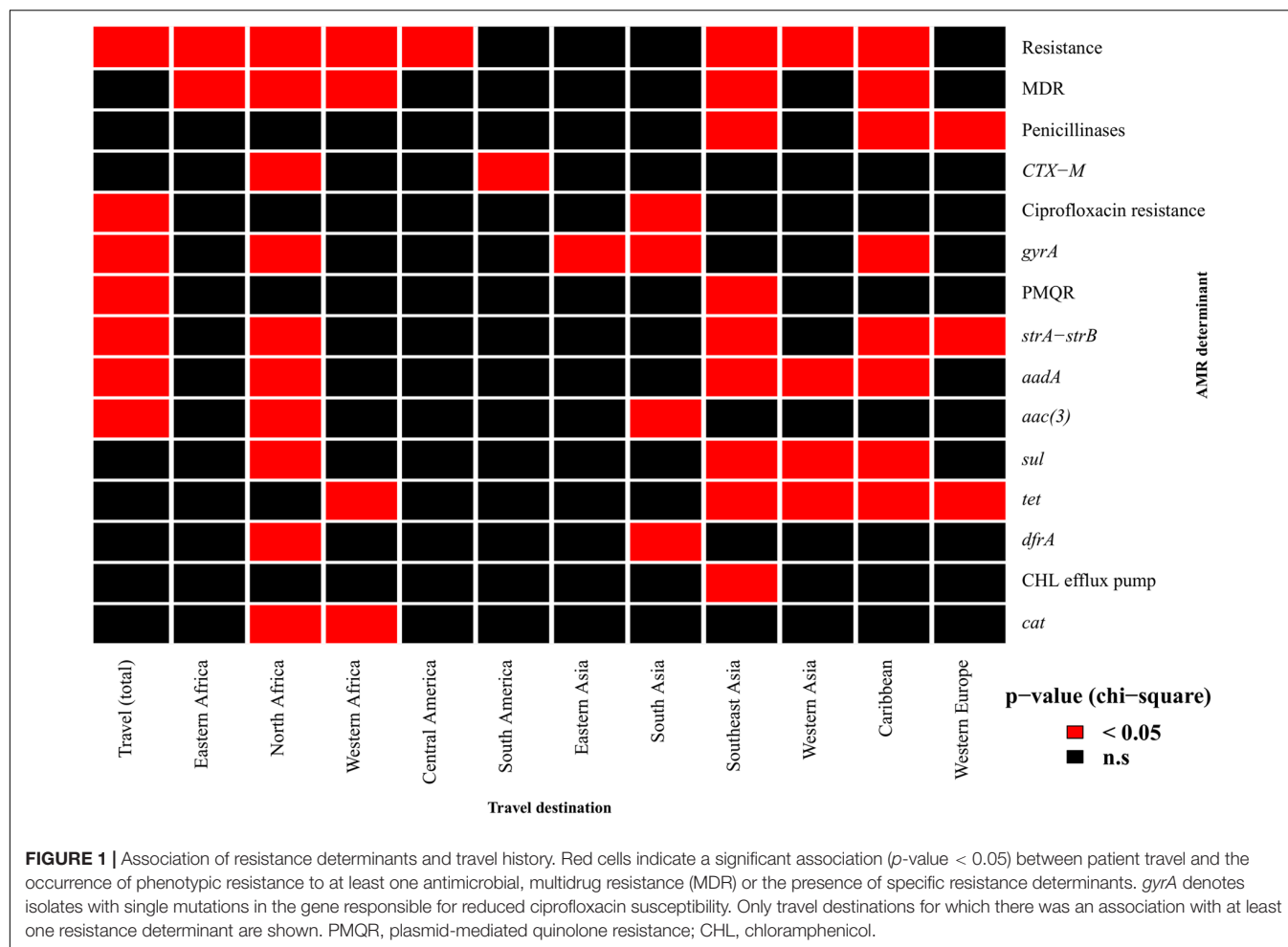
The implementation of WGS for surveillance of enteric pathogens has revolutionized the work of public health laboratories, as it allows inference of a multitude of pathogen characteristics in a single sequencing run, which would traditionally require a series of independent laboratory tests. A prominent example of the added value provided by WGS is the generation of AMR profiles from the sequences in real-time.

WGS has previously proven successful for prediction of AMR profiles in a variety of gastrointestinal pathogens, including *Shigella sonnei* (Sadouki et al., 2017), *Escherichia coli* (Stoesser et al., 2013; Tyson et al., 2015; Day et al., 2017a), *S. Typhi* (Day et al., 2017b) and smaller datasets of NTS (Zankari et al., 2013; Nair et al., 2016; McDermott et al., 2016). Our present comparison of phenotypic susceptibility testing and genotypic

prediction of AMR profiles based on WGS data for a much larger dataset, comprising 3,491 NTS isolates, identified 88 discordant results (0.17%) out of a possible 52,365 isolate/antimicrobial combinations, with the AMR profiles of 3,415 isolates (97.82%) completely matching for both approaches. Zankari et al. (2013) observed complete agreement of the two approaches for fifty *S. Typhimurium* isolates but only when excluding ciprofloxacin from the testing panel. Similar to our results, McDermott et al. (2016) found lower sensitivity and specificity for prediction of streptomycin resistance than for other antimicrobials tested.

Despite being an invaluable tool for surveillance purposes, AMR prediction based on WGS data is not yet deemed suitable to guide treatment choices (Ellington et al., 2017). Many MEs, where an isolate is phenotypically susceptible but carries genetic resistance determinants, seem to be associated with the breakpoints used for phenotypic testing. In some cases, the MICs are just below the recommended breakpoints but slight technical variations of the agar dilution method are possible so that the isolate would be falsely classified as susceptible. This seems to be an issue especially when testing for streptomycin resistance (Garcia-Migura et al., 2012), which would explain the relatively large number of mismatches in the present study. Recently, it has been suggested to adapt the breakpoint values to take into account MICs associated with the presence of specific resistance determinants (Tyson et al., 2017). Additionally, many of the resistance genes detected by the algorithm are plasmid-encoded but phenotypic susceptibility testing was carried out retrospectively. During storage and sub-culture of the isolates plasmids may be lost. Thus, genes detected during sequencing after initial cultivation might not be present when retrospective phenotypic testing is performed on a different colony. Furthermore, silent resistance genes, such as *bla*<sub>CMY-2</sub> and *tet* variants, have been observed previously in *Salmonella* (Heider et al., 2009; Adesiji et al., 2014). Other genes, such as the *aac*(6') variants, are normally silent and only become transcriptionally active in rare cases (Magnet et al., 1999).

The other mismatch category, the VMEs, where an isolate is genotypically predicted to be susceptible but exhibits phenotypic resistance, highlight the importance of active curation of the resistance gene database used for genotypic prediction. Mismatches are likely based on the presence of resistance determinants not included in the reference database used for prediction or on novel, unknown resistance mechanisms, the genetic determinants of which have not yet been described. Our pipeline, for instance, does not detect impermeability or efflux pump genes potentially contributing to ciprofloxacin resistance (Hopkins et al., 2005). Continuous scanning for new research findings should be carried out to enable identification of novel resistance mechanisms. These novel mechanisms will then be incorporated into the reference databases to maintain a high level of prediction sensitivity. Only recently, for example, computational methods identified previously unknown *qnr*-type fluoroquinolone resistance genes (Boulund et al., 2017). Despite these issues, the overall ME and VME rates of 0.13 and 0.04%, respectively, obtained in this study fall below the cut-offs of 3 and 1.5% from the US Food and Drug Administration for authorizing new susceptibility testing devices (FDA, 2009).



Specificity and sensitivity of ciprofloxacin resistance prediction exceeded 99% but we only considered isolates with an MIC > 0.5 mg/L for this evaluation. Traditionally, *gyrA* mutations in combination with *parC* mutations were thought to be required for ciprofloxacin resistance (Ruiz et al., 1997) and PMQRs on their own were not considered sufficient. Indeed, in our study, the majority of isolates showing resistance carried at least two mutations in both *gyrA* and *parC*. Thirty-seven had a PMQR gene, alone or in conjunction with a single *gyrA* mutation, which would normally be expected to result in reduced susceptibility instead of full resistance. Ciprofloxacin MICs for isolates carrying PMQR genes alone were found to range between 0.25 and 1 mg/L (Garcia-Fernandez et al., 2009) so that some isolates with this profile would be classed as resistant and some as having reduced susceptibility during phenotypic testing. Although the QRDR of *gyrA* is located between amino acids 67 and 106, mutations at positions 83 and 87 are most common (Yoshida et al., 1990). In our study, none of the isolates with mutations at other positions of the QRDR alone exhibited reduced ciprofloxacin susceptibility.

It has been suggested previously that an increased use of alternative antimicrobials, such as ciprofloxacin and extended-spectrum  $\beta$ -lactams, favored the re-emergence of susceptibility

to classical first-line drugs (Sood et al., 1999; Rahman et al., 2002). The limitation of our study was that it was biased toward serovars selected for their known high resistance rates, and therefore not a true representation of the expected serovar distribution in England and Wales over this time frame. We were therefore unable to assess changes in incidence of resistance to specific antimicrobials over the years. However, moving forward, genome-derived AMR profiling will provide a robust framework to explore longitudinal trends.

A worrying trend is the increase in resistance to extended-spectrum cephalosporins (Su et al., 2005; Mataseje et al., 2009). Since these antimicrobials are used as an alternative for treatment of invasive disease in case of resistance to ciprofloxacin, the emergence of co-resistance to both antimicrobial classes is of great concern. Co-resistance is especially prevalent in Asia (Lee et al., 2009) and was identified in thirteen isolates (0.37%) in this study, a slight increase from the 0.25% prevalence observed in the UK between 2010 and 2012 (Burke et al., 2014). Seven of the eighteen isolates carrying CTX-M-type ESBLs were associated with travel to Asia and a further seven with travel to North Africa. Similarly, as observed previously (Hopkins et al., 2007), PMQR genes were more likely to be found in isolates from

patients who had traveled to Asia. Extensively drug-resistant *S. Typhimurium*, like the one isolate in this study resistant to all eleven antimicrobial classes tested, have been found in Southeast Asia before (Benacer et al., 2010; Vo et al., 2010). Unfortunately, no travel history data was available for this isolate.

In addition to providing information on AMR for the entire NTS population, WGS-based prediction was able to highlight some interesting genotypic differences between the most common serovars *S. Typhimurium* and *S. Enteritidis* and the extensively drug-resistant *S. Kentucky*: ESBL genes and *aac(3)* variants, while found in *S. Typhimurium* and *S. Kentucky*, were absent in *S. Enteritidis*. Only *S. Kentucky* carried multiple mutations in the QRDRs of *gyrA* and *parC* but PMQR genes were less common than in other serovars. A more detailed investigation of these differences might lead to a better understanding of the varying outcomes associated with infections caused by different serovars of NTS (Jones et al., 2008).

## CONCLUSION

This large-scale study supports the suitability of WGS-based prediction to reliably replace phenotypic susceptibility testing for rapid monitoring of emerging trends in AMR patterns in NTS and for studying the spread of AMR genes in this pathogen population. Since sequencing is routinely used in public health laboratories already, it constitutes a time-saving alternative to traditional approaches that can further our understanding of resistance mechanisms as long as constant curation of the resistance gene database used is warranted. Prediction for further antimicrobials such as macrolides, fosfomycin and colistin will be validated in the near future to increase the robustness of the pipeline. Information derived from WGS-based studies can then be used to inform public health interventions aimed at limiting further dissemination of AMR genes and thus aid in the fight against the global AMR threat.

## REFERENCES

- Achtman, M., Wain, J., Weill, F. X., Nair, S., Zhou, Z., Sangal, V., et al. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 8:e1002776. doi: 10.1371/journal.ppat.1002776
- Adesiji, Y. O., Deekshit, V. K., and Karunasagar, I. (2014). Antimicrobial-resistant genes associated with *Salmonella* spp. isolated from human, poultry, and seafood sources. *Food Sci. Nutr.* 2, 436–442. doi: 10.1002/fsn3.119
- Ao, T. T., Feasey, N. A., Gordon, M. A., Keddy, K. H., Angulo, F. J., and Crump, J. A. (2015). Global burden of invasive nontyphoidal *Salmonella* disease, 2010<sup>1</sup>. *Emerg. Infect. Dis.* 21, 941–949. doi: 10.3201/eid2106.140999
- Ashton, P. M., Nair, S., Peters, T. M., Bale, J. A., Powell, D. G., Painset, A., et al. (2016). Identification of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ* 4:e1752. doi: 10.7717/peerj.1752
- Ashton, P. M., Owen, S. V., Kaindama, L., Rowe, W. P. M., Lane, C. R., Larkin, L., et al. (2017). Public health surveillance in the UK revolutionises our understanding of the invasive *Salmonella* Typhimurium epidemic in Africa. *Genome Med.* 9:92. doi: 10.1186/s13073-017-0480-7
- Benacer, D., Thong, K.-L., Watanabe, H., and Puthuchear, S. D. (2010). Characterization of drug resistant *Salmonella enterica* serotype Typhimurium

## AUTHOR CONTRIBUTIONS

SNa, EP, TD, PA, and GG conceived the study. SNe, MRD, MD, CJ, PA, KM, KH, NW, and TD contributed to the data analysis. SNe, CJ, and TD wrote the manuscript. All authors contributed to, read, and approved the final manuscript.

## FUNDING

This work was supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Gastrointestinal Infections at University of Liverpool in partnership with PHE, in collaboration with the University of East Anglia, University of Oxford, and the Quadram Institute. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00592/full#supplementary-material>

**TABLE S1 |** The genotypic resistance profile derived for each isolate in the study and the associated SRA accession of the deposited genome.

**TABLE S2 |** Resistance genes detected in non-typhoidal *Salmonella enterica*, their prevalence in all serovars, *S. Typhimurium*, *S. Enteritidis*, and *S. Kentucky* and the antimicrobial class to which they confer resistance.

**TABLE S3 |** Association of resistance determinants and travel history. Values in cells denote the number of isolates for which the travel history of the patient coincided with the presence of phenotypic resistance, multidrug resistance (MDR) or the presence of specific resistance determinants. *gyrA* denotes isolates with single mutations in the gene responsible for reduced ciprofloxacin susceptibility. Only travel destinations for which there was an association with at least one resistance determinant are shown. >CIP, ciprofloxacin resistance (MIC >0.5mg/L); PMQR, plasmid-mediated quinolone resistance; CHL, chloramphenicol.

- by antibiograms, plasmids, integrons, resistance genes and PFGE. *J. Microbiol. Biotechnol.* 20, 1042–1052. doi: 10.4014/jmb.0910.10028
- Boulund, F., Berglund, F., Flach, C.-F., Bengtsson-Palme, J., Marathe, N. P., Larsson, D. G. J., et al. (2017). Computational discovery and functional validation of novel fluoroquinolone resistance genes in public metagenomic data sets. *BMC Genomics* 18:682. doi: 10.1186/s12864-017-4064-0
- Boyd, D., Cloeckaert, A., Chaslus-Dancla, E., and Mulvey, M. R. (2002). Characterization of variant *Salmonella* genomic island 1 multidrug resistance regions from serovars Typhimurium DT104 and Agona. *Antimicrob. Agents Chemother.* 46, 1714–1722. doi: 10.1128/AAC.46.6.1714-1722.2002
- Burke, L., Hopkins, K. L., Meunier, D., Pinna, E., Fitzgerald-Hughes, D., Humphreys, H., et al. (2014). Resistance to third-generation cephalosporins in human non-typhoidal *Salmonella enterica* isolates from England and Wales, 2010–12. *J. Antimicrob. Chemother.* 69, 977–981. doi: 10.1093/jac/dkt469
- Day, M., Doumith, M., Jenkins, C., Dallman, T. J., Hopkins, K. L., Elson, R., et al. (2017b). Antimicrobial resistance in Shiga toxin-producing *Escherichia coli* serogroups O157 and O26 isolated from human cases of diarrhoeal disease in England, 2015. *J. Antimicrob. Chemother.* 72, 145–152. doi: 10.1093/jac/dkw371

- Day, M. R., Doumith, M., Do Nascimento, V., Nair, S., Ashton, P. M., Jenkins, C., et al. (2017a). Comparison of phenotypic and whole genome sequencing-derived antimicrobial resistance profiles of *Salmonella enterica* serovars Typhi and Paratyphi. *J. Antimicrob. Chemother.* 73, 365–372. doi: 10.1093/jac/dkx379
- Doumith, M., Godbole, G., Ashton, P., Larkin, L., Dallman, T., et al. (2016). Detection of the plasmid-mediated *mcr-1* gene conferring colistin resistance in human and food isolates of *Salmonella enterica* and *Escherichia coli* in England and Wales. *J. Antimicrob. Chemother.* 71, 2300–2305. doi: 10.1093/jac/dkw093
- EFSA (2017). The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2015. *EFSA J.* 15:212.
- Ellington, M. J., Ekelund, O., Aarestrup, F. M., Canton, R., Doumith, M., Giske, C., et al. (2017). The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin. Microbiol. Infect.* 23, 2–22. doi: 10.1016/j.cmi.2016.11.012
- FDA (2009). *Class II Special Controls Guidance Document: Antimicrobial Susceptibility Test (AST) Systems; Guidance for Industry and FDA*. Rockville, MD: US Food and Drug Administration, 42.
- Feasey, N. A., Hadfield, J., Keddy, K. H., Dallman, T. J., Jacobs, J., Deng, X., et al. (2016). Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat. Genet.* 48, 1211–1217. doi: 10.1038/ng.3644
- Garcia-Fernandez, A., Fortini, D., Veldman, K., Mevius, D., and Carattoli, A. (2009). Characterization of plasmids harbouring *qnrS1*, *qnrB2* and *qnrB19* genes in *Salmonella*. *J. Antimicrob. Chemother.* 63, 274–281. doi: 10.1093/jac/dkn470
- Garcia-Migura, L., Sunde, M., Karlsmose, S., Veldman, K., Schroeter, A., Guerra, B., et al. (2012). Establishing streptomycin epidemiological cut-off values for *Salmonella* and *Escherichia coli*. *Microb. Drug Resist.* 18, 88–93. doi: 10.1089/mdr.2011.0064
- Heider, L. C., Hoet, A. E., Wittum, T. E., Khaita, M. L., Love, B. C., Huston, C. L., et al. (2009). Genetic and phenotypic characterization of the *bla<sub>CMY</sub>* gene from *Escherichia coli* and *Salmonella enterica* isolated from food-producing animals, humans, the environment, and retail meat. *Foodborne Pathog. Dis.* 6, 1235–1240. doi: 10.1089/fpd.2009.0294
- Hopkins, K. L., Davies, R. H., and Threlfall, E. J. (2005). Mechanisms of quinolone resistance in *Escherichia coli* and *Salmonella*: recent developments. *Int. J. Antimicrob. Agents* 25, 358–373. doi: 10.1016/j.ijantimicag.2005.02.006
- Hopkins, K. L., Wootton, L., Day, M. R., and Threlfall, E. J. (2007). Plasmid-mediated quinolone resistance determinant *qnrS1* found in *Salmonella enterica* strains isolated in the UK. *J. Antimicrob. Chemother.* 59, 1071–1075. doi: 10.1093/jac/dkm081
- Jones, T. F., Ingram, L. A., Cieslak, P. R., Vugia, D. J., Tobin-D'angelo, M., Hurd, S., et al. (2008). Salmonellosis outcomes differ substantially by serotype. *J. Infect. Dis.* 198, 109–114. doi: 10.1086/588823
- Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., et al. (2009). Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res.* 19, 2279–2287. doi: 10.1101/gr.091017.109
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Langridge, G. C., Fookes, M., Connor, T. R., Feltwell, T., Feasey, N., Parsons, B. N., et al. (2015). Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 863–868. doi: 10.1073/pnas.1416707112
- Lee, H.-Y., Su, L.-H., Tsai, M.-H., Kim, S.-W., Chang, H.-H., Jung, S.-I., et al. (2009). High rate of reduced susceptibility to ciprofloxacin and ceftriaxone among nontyphoid *Salmonella* clinical isolates in Asia. *Antimicrob. Agents Chemother.* 53, 2696–2699. doi: 10.1128/AAC.01297-08
- Le Hello, S., Harrois, D., Bouchrif, B., Sontag, L., Elhani, D., Guibert, V., et al. (2013). Highly drug-resistant *Salmonella enterica* serotype Kentucky St198-X1: a microbiological study. *Lancet Infect. Dis.* 13, 672–679. doi: 10.1016/S1473-3099(13)70124-5
- Magnet, S., Courvalin, P., and Lambert, T. (1999). Activation of the cryptic *aac(6')*-Iy aminoglycoside resistance gene of *Salmonella* by a chromosomal deletion generating a transcriptional fusion. *J. Bacteriol.* 181, 6650–6655.
- Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., et al. (2010). The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin. Infect. Dis.* 50, 882–889. doi: 10.1086/650733
- Mataseje, L. F., Xiao, J., Kost, S., Ng, L. K., Dore, K., Mulvey, M. R., et al. (2009). Characterization of Canadian cefoxitin-resistant non-typhoidal *Salmonella* isolates, 2005–06. *J. Antimicrob. Chemother.* 64, 723–730. doi: 10.1093/jac/dkp249
- McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., et al. (2016). Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. *Antimicrob. Agents Chemother.* 60, 5515–5520. doi: 10.1128/AAC.01030-16
- Nair, S., Ashton, P., Doumith, M., Connell, S., Painset, A., Mwaigwisya, S., et al. (2016). WGS for surveillance of antimicrobial resistance: a pilot study to detect the prevalence and mechanism of resistance to azithromycin in a UK population of non-typhoidal *Salmonella*. *J. Antimicrob. Chemother.* 71, 3400–3408. doi: 10.1093/jac/dkw318
- Parry, C. M., Thomas, S., Aspinall, E. J., Cooke, R. P., Rogerson, S. J., Harries, A. D., et al. (2013). A retrospective study of secondary bacteraemia in hospitalised adults with community acquired non-typhoidal *Salmonella* gastroenteritis. *BMC Infect. Dis.* 13:107. doi: 10.1186/1471-2334-13-107
- Rahman, M., Ahmad, A., and Shoma, S. (2002). Decline in epidemic of multidrug resistant *Salmonella* Typhi is not associated with increased incidence of antibiotic-susceptible strain in Bangladesh. *Epidemiol. Infect.* 129, 29–34. doi: 10.1017/S0950268802007203
- Ruiz, J., Castro, D., Goni, P., Santamaria, J. A., Borrego, J. J., and Vila, J. (1997). Analysis of the mechanism of quinolone resistance in nalidixic acid-resistant clinical isolates of *Salmonella* serotype Typhimurium. *J. Med. Microbiol.* 46, 623–628. doi: 10.1099/00222615-46-7-623
- Sadouki, Z., Day, M. R., Doumith, M., Chattaway, M. A., Dallman, T. J., Hopkins, K. L., et al. (2017). Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Shigella sonnei* isolated from cases of diarrhoeal disease in England and Wales, 2015. *J. Antimicrob. Chemother.* 72, 2496–2502. doi: 10.1093/jac/dkx170
- Sood, S., Kapil, A., Das, B., Jain, Y., and Kabra, S. K. (1999). Re-emergence of chloramphenicol-sensitive *Salmonella typhi*. *Lancet*, 353, 1241–1242. doi: 10.1016/S0140-6736(99)00637-6
- Stoesser, N., Batty, E. M., Eyre, D. W., Morgan, M., Wyllie, D. H., Del Ojo Elias, C., et al. (2013). Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J. Antimicrob. Chemother.* 68, 2234–2244. doi: 10.1093/jac/dkt180
- Su, L. H., Wu, T. L., Chia, J. H., Chu, C., Kuo, A. J., and Chiu, C. H. (2005). Increasing ceftriaxone resistance in *Salmonella* isolates from a university hospital in Taiwan. *J. Antimicrob. Chemother.* 55, 846–852. doi: 10.1093/jac/dki116
- Tam, C. C., O'Brien, S. J., Tompkins, D. S., Bolton, F. J., Berry, L., Dodds, J., et al. (2012). Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease. *Clin. Infect. Dis.* 54, 1275–1286. doi: 10.1093/cid/cis028
- Tassios, P. T., Gazouli, M., Tzelepi, E., Milch, H., Kozlova, N., Sidorenko, S., et al. (1999). Spread of a *Salmonella typhimurium* clone resistant to expanded-spectrum cephalosporins in three European countries. *J. Clin. Microbiol.* 37, 3774–3777.
- Threlfall, E. J., Frost, J. A., Ward, L. R., and Rowe, B. (1996). Increasing spectrum of resistance in multiresistant *Salmonella typhimurium*. *Lancet* 347, 1053–1054. doi: 10.1016/S0140-6736(96)90199-3
- Threlfall, E. J., Ward, L. R., and Rowe, B. (1997). Increasing incidence of resistance to trimethoprim and ciprofloxacin in epidemic *Salmonella typhimurium* DT104 in England and Wales. *Euro Surveill.* 2, 81–84. doi: 10.2807/esm.02.11.00187-en
- Tyson, G. H., McDermott, P. F., Li, C., Chen, Y., Tadesse, D. A., Mukherjee, S., et al. (2015). WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *J. Antimicrob. Chemother.* 70, 2763–2769. doi: 10.1093/jac/dkv186
- Tyson, G. H., Zhao, S., Li, C., Ayers, S., Sabo, J. L., Lam, C., et al. (2017). Establishing genotypic cutoff values to measure antimicrobial resistance in *Salmonella*. *Antimicrob. Agents Chemother.* 61:e02140-16. doi: 10.1128/AAC.02140-16
- Villa, L., Guerra, B., Schmoger, S., Fischer, J., Helmuth, R., Zong, Z., et al. (2015). IncA/C plasmid carrying *bla<sub>NDM-1</sub>*, *bla<sub>CMY-16</sub>*, and *fosA3* in a *Salmonella*



- enterica serovar corvallis strain isolated from a migratory wild bird in Germany. *Antimicrob. Agents Chemother.* 59, 6597–6600. doi: 10.1128/AAC.00944-15
- Villa, L., Mammina, C., Miriagou, V., Tzouveleakis, L. S., Tassios, P. T., Nastasi, A., et al. (2002). Multidrug and broad-spectrum cephalosporin resistance among *Salmonella enterica* serotype enteritidis clinical isolates in Southern Italy. *J. Clin. Microbiol.* 40, 2662–2665. doi: 10.1128/JCM.40.7.2662-2665.2002
- Vo, A. T. T., Van Duijkeren, E., Gaastra, W., and Fluit, A. C. (2010). Antimicrobial resistance, class 1 integrons, and genomic island 1 in *Salmonella* isolates from Vietnam. *PLoS One* 5:e9440. doi: 10.1371/journal.pone.0009440
- Wain, J., Hendriksen, R. S., Mikoleit, M. L., Keddy, K. H., and Ochiai, R. L. (2015). Typhoid fever. *Lancet* 385, 1136–1145. doi: 10.1016/S0140-6736(13)62708-7
- Yoshida, H., Bogaki, M., Nakamura, M., and Nakamura, S. (1990). Quinolone resistance-determining region in the DNA gyrase *gyrA* gene of *Escherichia coli*. *Antimicrob. Agents Chemother.* 34, 1271–1272. doi: 10.1128/AAC.34.6.1271
- Zankari, E., Hasman, H., Kaas, R. S., Seyfarth, A. M., Agersø, Y., Lund, O., et al. (2013). Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J. Antimicrob. Chemother.* 68, 771–777. doi: 10.1093/jac/dks496
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Nuert, Nair, Day, Doumith, Ashton, Mellor, Jenkins, Hopkins, Woodford, de Pinna, Godbole and Dallman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Antibiotic-Induced Alterations in Gut Microbiota Are Associated with Changes in Glucose Metabolism in Healthy Mice

Richard R. Rodrigues<sup>1</sup>, Renee L. Greer<sup>2</sup>, Xiaoxi Dong<sup>1</sup>, Karen N. DSouza<sup>1</sup>, Manoj Gurung<sup>2</sup>, Jia Y. Wu<sup>2</sup>, Andrey Morgun<sup>1\*</sup> and Natalia Shulzhenko<sup>2\*</sup>

<sup>1</sup> Department of Pharmaceutical Sciences, Oregon State University, Corvallis, OR, United States, <sup>2</sup> Department of Biomedical Sciences, Oregon State University, Corvallis, OR, United States

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Amanda Ellen Ramer-Tait,  
University of Nebraska–Lincoln,  
United States  
Jonathan Badger,  
National Cancer Institute (NIH),  
United States

### \*Correspondence:

Natalia Shulzhenko  
natalia.shulzhenko@oregonstate.edu  
Andrey Morgun  
andriy.morgun@oregonstate.edu

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 18 August 2017

**Accepted:** 08 November 2017

**Published:** 22 November 2017

### Citation:

Rodrigues RR, Greer RL, Dong X,  
DSouza KN, Gurung M, Wu JY,  
Morgun A and Shulzhenko N (2017)  
Antibiotic-Induced Alterations in Gut  
Microbiota Are Associated with  
Changes in Glucose Metabolism  
in Healthy Mice.  
Front. Microbiol. 8:2306.  
doi: 10.3389/fmicb.2017.02306

The gut microbiome plays an important role in health and disease. Antibiotics are known to alter gut microbiota, yet their effects on glucose tolerance in lean, normoglycemic mice have not been widely investigated. In this study, we aimed to explore mechanisms by which treatment of lean mice with antibiotics (ampicillin, metronidazole, neomycin, vancomycin, or their cocktail) influences the microbiome and glucose metabolism. Specifically, we sought to: (i) study the effects on body weight, fasting glucose, glucose tolerance, and fasting insulin, (ii) examine the changes in expression of key genes of the bile acid and glucose metabolic pathways in the liver and ileum, (iii) identify the shifts in the cecal microbiota, and (iv) infer interactions between gene expression, microbiome, and the metabolic parameters. Treatment with individual or a cocktail of antibiotics reduced fasting glucose but did not affect body weight. Glucose tolerance changed upon treatment with cocktail, ampicillin, or vancomycin as indicated by reduced area under the curve of the glucose tolerance test. Antibiotic treatment changed gene expression in the ileum and liver, and shifted the alpha and beta diversities of gut microbiota. Network analyses revealed associations between *Akkermansia muciniphila* with fasting glucose and liver farnesoid X receptor (Fxr) in the top ranked host-microbial interactions, suggesting possible mechanisms by which this bacterium can mediate systemic changes in glucose metabolism. We observed *Bacteroides uniformis* to be positively and negatively correlated with hepatic Fxr and Glucose 6-phosphatase, respectively. Overall, our transkingdom network approach is a useful hypothesis generating strategy that offers insights into mechanisms by which antibiotics can regulate glucose tolerance in non-obese healthy animals. Experimental validation of our predicted microbe-phenotype interactions can help identify mechanisms by which antibiotics affect host phenotypes and gut microbiota.

**Keywords:** antibiotics, gut microbiota, glucose tolerance, lean, non-obese, transkingdom networks

**Abbreviations:** AUC, area under the curve; Fgf15, fibroblast growth factor 15; Fxr, farnesoid x receptor; G6pase, glucose 6-phosphatase; Glut1, glucose transporter 1; GTT, glucose tolerance test; Hk1, hexokinase 1; Hk2, hexokinase 2; Insr, insulin receptor; Pck1, phosphoenolpyruvate carboxykinase 1; Shp, small heterodimer partner; Tgr5, G protein-coupled bile acid receptor 1.

## INTRODUCTION

The human gastrointestinal tract contains a multitude of microbiota, including bacteria, viruses, and fungi (Utzschneider et al., 2016). Their genome, although variable between individuals (Human Microbiome Project Consortium, 2012), is capable of a diverse set of functions that may influence the host's metabolic and immune systems (Tremaroli and Backhed, 2012; Greer et al., 2013; Sanz et al., 2015), including normal homeostasis (Utzschneider et al., 2016). Changes in the gut microbes have recently been associated with various diseases (Qin et al., 2012; Karlsson et al., 2013; Wu et al., 2015). For example, changes in *Lactobacillus*, *Clostridium*, *Ruminococcus* sp., *E. coli*, *Bacteroides*, *Akkermansia muciniphila* are observed in diabetic and obese patients (Qin et al., 2012; Karlsson et al., 2013; Murri et al., 2013; Chakraborti, 2015; Kasai et al., 2015; Sanz et al., 2015). These diverse results indicate a need for a better understanding of the mechanistic roles specific taxa play in the regulation of host metabolic functions.

Antibiotics add an interesting dynamic to the host-microbiome relationship. Although, antibiotics are well-known to cause short (Perez-Cobas et al., 2013; Pallav et al., 2014; Panda et al., 2014) and long-term (De La Cochetiere et al., 2005; Jernberg et al., 2007; Dethlefsen et al., 2008; Jakobsson et al., 2010; Dethlefsen and Relman, 2011; Raymond et al., 2016) alterations in the gut microbiome, there is a lack of consensus on their effects on glucose tolerance, body weight and other metabolic parameters (Francino, 2015; Mikkelsen et al., 2016). Moreover, effects of antibiotics in lean, normoglycemic mice as compared to mouse obesity models have not been widely investigated. An intervention study in healthy, glucose tolerant young human males treated with 4-days broad-spectrum antibiotics cocktail showed shifts in the cultivable gut microbiota but no changes in postprandial plasma glucose and serum insulin (Mikkelsen et al., 2016). Due to the use of a broad-acting antibiotic cocktail in a short course as well as the use of fecal samples for culture-based bacterial assessment, this study provides limited insight on a comprehensive picture of changes in intestinal microbes and on associations between individual antibiotics and specific intestinal microbes. Understanding antibiotic-microbiome interactions and their effects on glucose metabolism in healthy mammals is critical for identifying initial changes in microbiota that eventually may lead to diseases such as obesity and diabetes.

In this study, we aimed to understand the regulatory mechanisms by which individual antibiotics and their cocktail influence the cecal microbiome and host phenotypes in lean mice, namely, gene expression and metabolic parameters. By treating lean mice with different antibiotics we sought to: (i) study the effects on body weight, fasting glucose, glucose tolerance, and fasting insulin, (ii) examine the changes in expression of key genes of the bile acid and glucose metabolic pathways in the liver and ileum, (iii) identify the shifts in the cecal microbiota, and (iv) infer interactions between gene expression, microbiome, and the metabolic parameters. We repeated the entire experiment twice and

performed meta-analyses to increase the confidence of our results.

## MATERIALS AND METHODS

### Mice and Antibiotics Treatment

Eight weeks old adult male Swiss Webster mice were initially purchased from Taconic Biosciences (Germantown, MD, United States). Mice were housed at the Laboratory Animal Resource Center at Oregon State University for 3–5 days for acclimation under standard 12-h light cycle with free access to food (5001, Research Diets) and water. Experimental procedures were carried out in accordance with protocols approved by the Oregon State University Institutional Animal Care and Use Committee. Mice were given single, cocktail, or no antibiotics for 4 weeks to create a stable altered microbiome. Antibiotics were administered in autoclaved drinking water individually, or in a cocktail for 4 weeks in the following concentrations: ampicillin ( $1 \text{ g l}^{-1}$ ), metronidazole ( $1 \text{ g l}^{-1}$ ), neomycin trisulfate ( $1 \text{ g l}^{-1}$ ), and vancomycin ( $0.5 \text{ g l}^{-1}$ ). This time course is consistent with standard antibiotic administration used in multiple studies for altering microbiota (Rakoff-Nahoum et al., 2004; Morgun et al., 2015; Greer R.L. et al., 2016). Each group consisted of five mice per experiment, total 30 mice per experiment, except for four mice in the cocktail group from the second experiment. Water consumption was monitored over the 4 weeks treatment period and all groups showed consumption equivalent to control water.

### Glucose Tolerance Testing

Mice were fasted for 6 h during the light phase with free access to water. A concentration of  $2 \text{ mg kg}^{-1}$  glucose (Sigma-Aldrich) was injected intraperitoneally. Blood glucose was measured at 0 (immediately before glucose injection), 15, 30, 60, and 120 min with a Freestyle Lite glucometer (Abbot Diabetes Care).

### Serum Collection and Hormone Measures

Mice were fasted for 6 h during the light phase with free access to water. Serum was collected via submandibular bleed using BD microtainer serum separator tubes. Fasting insulin was measured by ultrasensitive ELISA (Crystal Chem) according to manufacturer's protocols.

### Bacterial DNA Extraction, 16S rRNA Gene Library Preparation and PCR

Unflushed cecal tissue and content was suspended in 1.4 ml ASL buffer (Qiagen) and homogenized with 2.8 mm ceramic beads followed by 0.5 mm glass beads using an OMNI Bead Ruptor (OMNI International). DNA was extracted from the entire resulting suspension using QIAamp DNA Stool Mini Kit (Qiagen) according to manufacturer's protocol. DNA was quantified using Qubit broad range DNA assay (Life Technologies). The V4 region of 16s rRNA gene was amplified using universal primers (515f and 806r) (Caporaso et al., 2012). Individual samples were barcoded, pooled to construct the

sequencing library, and then sequenced using an Illumina Miseq (Illumina, San Diego, CA, United States) to generate pair-ended 250 nt reads. Quantitative PCR was performed for *A. muciniphila* as described in Schneeberger et al. (2015) with DNA for standard curve isolated from the cultivated microbe.

## RNA Preparation and Gene Expression Analysis

Liver and ileum (flushed out of content) were collected and snap frozen prior to RNA extraction. Liver was homogenized using OMNI Rotor-Stator Homogenizer in Trizol and RNA was extracted using Trizol/chloroform extraction followed by the RNeasy Mini kit (Qiagen). Ileum RNA was extracted using OMNI Bead Ruptor and 2.8 mm ceramic beads (OMNI International) in RLT buffer followed by Qiashredder and RNeasy kit using Qiacube (Qiagen) automated extraction according to manufacturer's specifications. Total RNA was quantified using Nanodrop (Thermo Scientific). Complementary DNA was prepared using iScript reverse transcription kit (Bio-Rad) and qPCR was performed using QuantiFast SYBR mix (Qiagen) and StepOne Plus Real Time PCR system and software (Applied Biosystems). Primers used for qPCR are listed in **Supplementary Table S1**.

## Statistical Analysis of Phenotypic Data

An outlier value per group per experiment was removed (if  $p$ -value < 5%) for each phenotype (metabolic parameters and genes) using the default Grubb's test from R package *outliers* v0.14 (Komsta, 2011). The data was  $\log_2$  transformed and differential phenotypes (antibiotics vs. control) were detected using *limma* (Ritchie et al., 2015) (Bioconductor 3.4, BiocInstaller 1.24.0, R 3.3.2) per experiment. A combined Fisher's  $p$ -value was calculated for each phenotype from the  $p$ -values for the *limma*  $t$ -statistic from each experiment. A false discovery rate (FDR) was calculated on the combined  $p$ -values. Change in phenotype was considered statistically significant if the phenotype had same direction of (abx/control) fold change in both experiments, individual  $p$ -value < 20% in each experiment, Fisher's combined  $p$ -value (Fisher, 1932) < 5% and FDR < 10%. The dot plots for the phenotypes were generated using R package *ggplot2* (Wickham, 2009) and the GTT curves were generated using GraphPad Prism software v7.03.

## Analyses of 16S rRNA Gene Sequencing Data

The samples were demultiplexed and forward-end fastq files were analyzed using QIIME v. 1.9.1 (Caporaso et al., 2010). The default quality filter parameters from QIIME's *split\_libraries\_fastq.py* were applied to retain high quality reads (Phred quality score  $\geq 20$  and minimum read length = 75% of 250 nucleotides). A closed reference OTU picking with 97% sequence similarity was performed using UCLUST (Edgar, 2010) and Greengenes reference database v13.8 (DeSantis et al., 2006; McDonald et al., 2012) to cluster 16S rRNA gene sequence reads into OTUs and assign taxonomy. The

reference sequence of an OTU from the Greengenes database was used to obtain species level taxonomic assignment using Megablast (Altschul et al., 1997; Morgulis et al., 2008) (top hit using default parameters). A threshold of 99% cumulative abundance across all samples in an experiment was used to retain abundant microbes, thus removing OTUs with approximately <0.01% abundance across all samples in that experiment. The read counts were normalized using cumulative sum scaling (Paulson et al., 2013) followed by quantile normalization.

The normalized OTU tables were used for diversity and statistical analysis. Briefly, a sampling depth of 200,000 sequences per sample was used for rarefaction. The alpha diversity metrics were calculated on unrarefied and rarefied OTU tables (**Supplementary Table S2**). The Shannon diversity index (from rarefied data) for samples with and without antibiotics treatment was compared with a non-parametric  $t$ -test. The difference was considered to be statistically significant if the direction of (abx/control) fold change in both experiments is the same, individual  $p$ -value < 2% in each experiment, Fisher's combined  $p$ -value < 0.1% and FDR < 0.1%. Beta diversity was calculated using weighted UniFrac (Lozupone and Knight, 2005) and the distances were used for PCoA (Gower, 1998) and visualized using EMPERor (Vazquez-Baeza et al., 2013). The taxonomic summary bar plots were used to visualize abundance at the phylum and order levels.

The  $\log_2$  transformed OTU tables were used for limma analysis. Meta-analysis was performed using the same criteria as applied for phenotypes to identify differentially abundant OTUs. A heatmap with row scaling was generated for each experiment using R packages *ggfortify* v0.2 (Horikoshi and Tang, 2016) and *gplots* v3.0.1 (Warnes et al., 2016). Hierarchical clustering was used to group OTUs (rows) based on similar abundance patterns across the groups in the first experiment and the same row order was used for the second experiment without row-wise clustering.

## Network Reconstruction and Prioritizing Microbe-Phenotype Edges

Spearman rank correlations were calculated between all pairs of genes, microbes, and metabolic parameters across all samples or per-group in an experiment. A combined Fisher's  $p$ -value was calculated for each pair from the  $p$ -values for the correlation from each experiment. A FDR was calculated on the combined  $p$ -values separately for the following correlations: (i) within genes, (ii) within metabolic parameters, (iii) between genes and metabolic parameters, and (iv) between OTUs and phenotypes (genes or metabolic parameters).

We retained edges that satisfy the following criteria: the sign of correlation coefficients in the two experiments should be consistent, individual  $p$ -value of correlation within each experiment is <20%, combined Fisher's  $p$ -value of all experiments <5% and FDR cutoff of 10% for edges without a microbial node (i, ii, and iii), whereas 1% for edges containing at least one microbial node (iv).

Next, the transkingdom network was generated (Dong et al., 2015; Morgun et al., 2015; Greer R.L. et al., 2016; Rodrigues



et al., 2017) by keeping the criteria-satisfying phenotypic (i, ii, and iii) and OTU-phenotype (iv) edges, where the OTU has >0.5% median abundance across the two experiments in at least one group.

Finally, an OTU-phenotype edge was retained if it showed consistent sign of per-group Spearman correlation coefficient between the two experiments, principles of causality compliancy (Yambartsev et al., 2016) [i.e., satisfied fold change relationship between the two partners in the appropriate (abx vs. control) comparison] in at least one group, and the same sign of correlation coefficient across different groups. To put this bipartite network in perspective of the phenotypic connections a phenotypic edge was included (only during visualization) if its strength of correlation was stronger than at least one OTU-phenotype edges connecting the phenotypes. Network topology statistics, namely degree and betweenness centrality (BC), were calculated using NetworkAnalyzer (Assenov et al., 2008) in Cytoscape v3.5 (Shannon et al., 2003). These edges were ranked using a score of maximum (per-group OTU abundance)  $\times$  absolute [median (per-group correlation)] to prioritize OTUs and the phenotypes they potentially affect, where the per-group OTU abundance and correlation are medians across the two experiments. The top hit of BLAST for the Greengenes representative sequence for an OTU was used to obtain species level identification.

## Data Availability

Raw reads of 16S rRNA gene sequencing have been deposited at NCBI under BioProject PRJNA394608, Biosamples of SAMN07356206 – SAMN07356264, Sequence Read Archive SRP112596.

## RESULTS

Lean, normoglycemic male mice were left untreated, or were treated with ampicillin, metronidazole, neomycin or vancomycin, or a cocktail containing all four antibiotics for 4 weeks to study the effects of antibiotic treatment on glucose tolerance, genes involved in glucose and bile acid metabolism, and the gut microbiota. Antibiotics resulted in different patterns of changes in the metabolic parameters, gene expression, and intestinal microbiome.

### Antibiotics Improved Glucose Tolerance in Lean Mice

No metabolic parameter worsened following antibiotics treatment (Figure 1 and Supplementary Figure S1). We observed that treatment with individual or cocktail of antibiotics reduced fasting glucose, but did not change body weight. Glucose tolerance improved upon treatment with cocktail, ampicillin, or vancomycin as indicated by reduced AUC of the GTT. Treatment with all antibiotics, including metronidazole or neomycin reduced fasting glucose levels, however, the latter two did not cause changes in systemic glucose tolerance. Fasting insulin was reduced only when the mice were treated with

vancomycin. Overall, glucose metabolism was regulated by antibiotic treatment.

### Antibiotics Changed Expression of Genes Involved in Glucose and Bile Acid Metabolism

Tissue specific host gene expression is important in many metabolic processes (Thomas et al., 2008; Chiang, 2013) and regulated by gut microbiota (Larsson et al., 2012). These, along with the knowledge that intestinal glucose metabolism can control systemic glucose levels (Saeidi et al., 2013), led us to examine the expression of key glucose and bile acid metabolic genes in the liver and the ileum.

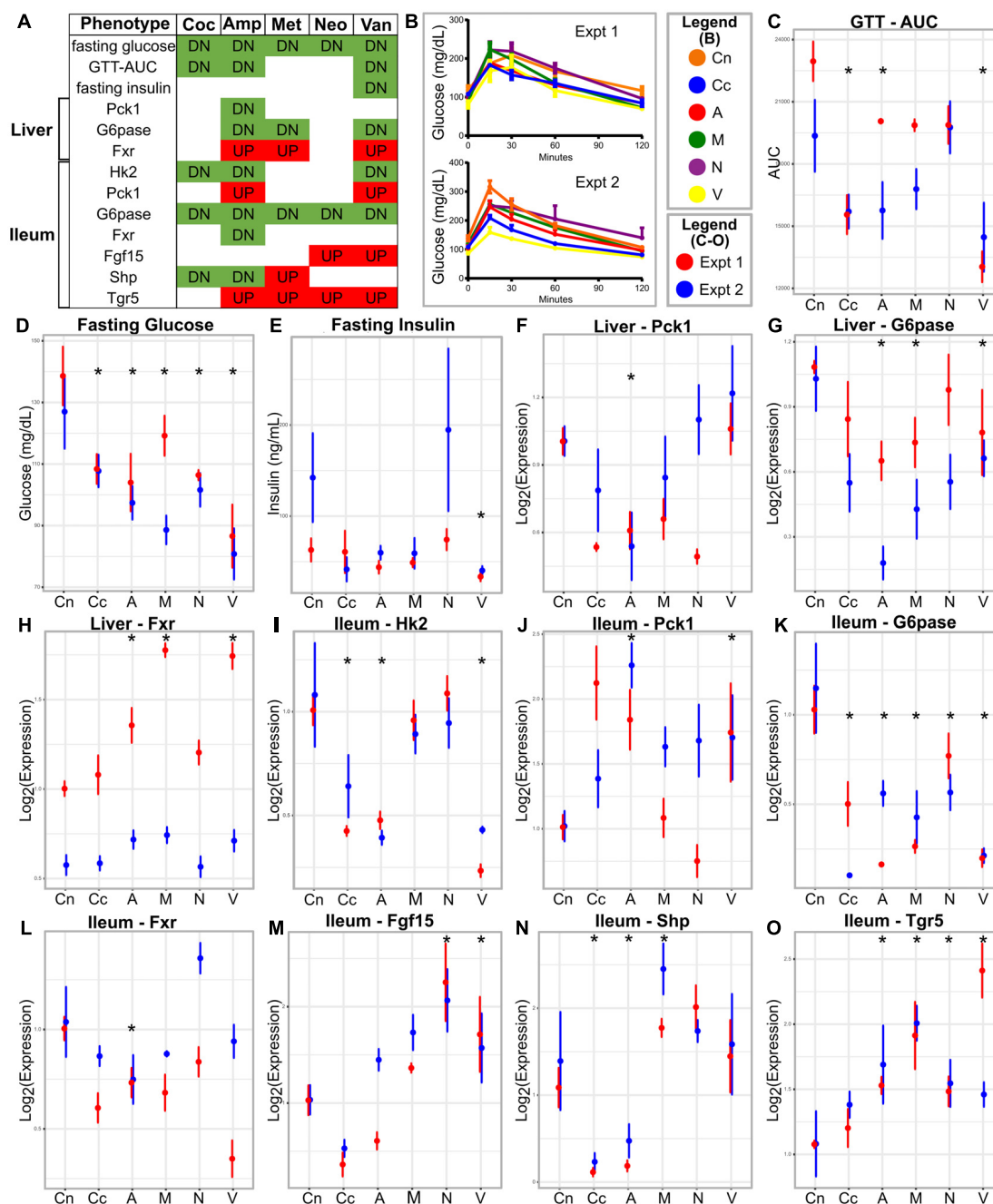
The majority of the tested genes in the ileum showed changes in expression due to antibiotic treatment (Figure 1). Ileum Hk2 and G6pase transcripts showed decreased expressions after treatment with cocktail, ampicillin, or vancomycin. Ileum Pck1 and Tgr5 mRNA were increased after treatment with ampicillin or vancomycin, but showed no changes after with cocktail. Ileal Hk1 and Glut1 did not change gene expression after antibiotics, whereas, Fgf15, Fxr, and Shp showed antibiotic-specific patterns in expression.

Only three genes showed differential expression in the liver following antibiotic treatment (Figure 1). Fxr and G6pase showed increased and decreased expression, respectively, in ampicillin or vancomycin treated mice. Pck1 showed lower expressions in ampicillin treated samples. Hk2 and Insr genes in the liver did not change following antibiotics treatment.

Despite some variability in tissue specific behavior of genes in response to antibiotics, the improved glucose tolerance upon antibiotic treatment suggests that relationships between gene expression and metabolic parameters are mostly preserved across all groups. Hence, we constructed a correlation network consisting of (differentially expressed) genes and (differentially abundant) metabolic parameters using all samples per experiment (Figure 2). Genes from the ileum, including G6pase, Hk2, and Fxr were strongly connected with the GTT-AUC. The Fxr gene in the liver was positively correlated with the ileum Tgr5 but negatively correlated with ileum Fxr and with fasting glucose and GTT. Altogether, this network indicates opposite effects of intestinal and liver Fxr on glucose metabolism. Furthermore, it also suggests that increased glycolytic gene expression program in ileum is connected to worsening of systemic glucose metabolism.

### Antibiotics Caused Shifts in Microbial Communities

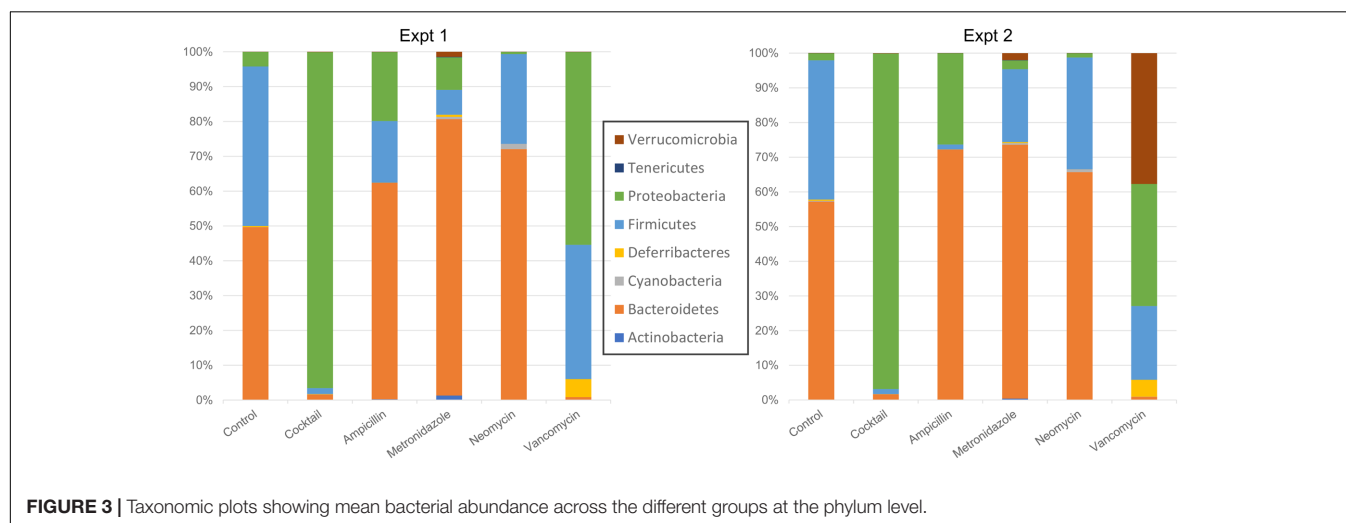
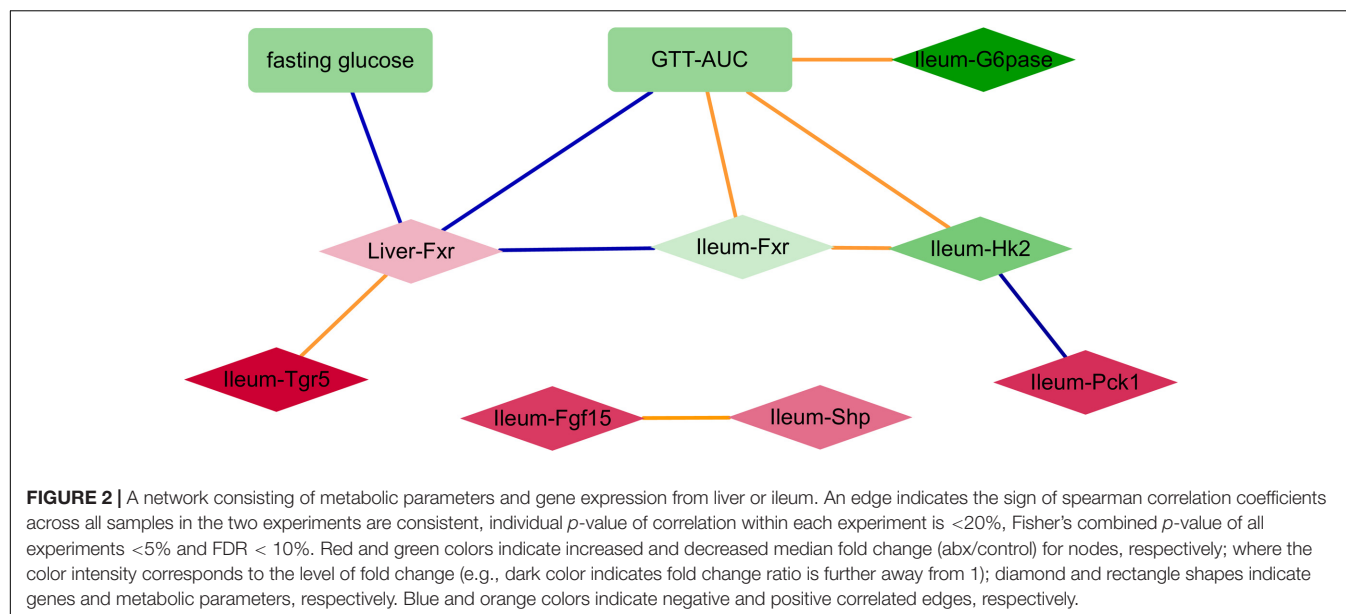
Microbiome composition is known to be affected by antibiotics (De La Cochetiere et al., 2005; Jernberg et al., 2007; Dethlefsen et al., 2008; Jakobsson et al., 2010; Dethlefsen and Relman, 2011; Perez-Cobas et al., 2013; Pallav et al., 2014; Panda et al., 2014; Raymond et al., 2016) and involved in metabolic processes (Larsson et al., 2012; Tremaroli and Backhed, 2012; Sanz et al., 2015; Utzschneider et al., 2016), so we hypothesized that gut microbes might play a mechanistic role in the effect of antibiotics (Morgun et al., 2015; Greer R. et al., 2016; Greer



**FIGURE 1 |** Metabolic parameters and gene expression in antibiotic-treated and control animals. **(A)** Summary table; the red and green colors indicate increase and decrease, respectively, in antibiotic treated group compared to the control. **(B)** GTT curves for the antibiotics treated and control groups in the two experiments. **(C–E)** Metabolic parameters and gene **(F–O)** expression represented as means with standard error bars. The red and blue colors indicate experiments one and two, respectively. Asterisks indicate parameters that show statistically significant differences upon antibiotics treatment compared to untreated control mice [same direction of (abx/control) fold change in both experiments, individual  $p$ -value < 20% in each experiment, Fisher's combined  $p$ -value < 5% and FDR < 10%]. Cn, Control; Coc or Cc, cocktail; Amp or A, ampicillin; Met or M, metronidazole; Neo or N, neomycin; Van or V, vancomycin.

R.L. et al., 2016) on host glucose metabolism (Caesar et al., 2012; Greer R.L. et al., 2016). Sequencing the 16S rRNA gene of the cecal microbiome from the two experiments provided a total of 14,321,948 high quality reads with mean length of 248.50 bases and standard deviation of 9.42. A threshold of

99% cumulative abundance across all samples per experiment retained 734 and 677 OTUs in the two experiments (overlap of 561 OTUs) with 5,450,867 and 5,525,927 assigned sequences to the OTUs. The alpha diversity metrics on the normalized and rarefied OTU tables are provided in **Supplementary**

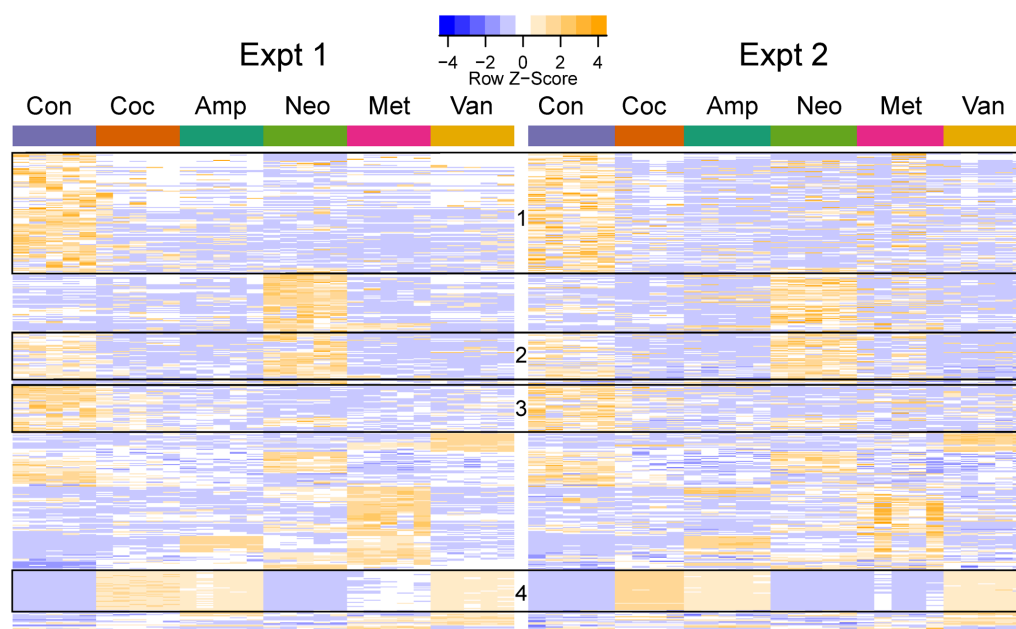


**Table S2.** As expected, the cocktail of antibiotics reduced the diversity of the samples compared to untreated or individual antibiotics (**Supplementary Figures S2, S3**). Shannon diversity comparisons showed that alpha diversity decreased when treated with cocktail, ampicillin, or vancomycin (**Supplementary Figure S2**).

A PCoA analysis using the weighted UniFrac suggested that the overall community composition from vancomycin and ampicillin treatment was closer to that when treated with antibiotics cocktail (**Supplementary Figure S3**). At the phylum level, Firmicutes decreased in cocktail, ampicillin, and metronidazole treated samples. Bacteroidetes decreased upon cocktail and vancomycin treatment but increased when treated with metronidazole (**Figures 3, 4** and **Supplementary Figure S5**; Sheet in **Supplementary Table S3**). The treatment with antibiotics showed similar patterns of change in the abundant bacteria at the order level, while less abundant bacteria showed

antibiotic specific changes (**Supplementary Figures S4, S6**). Vancomycin treatment increased Verrucomicrobiales in both experiments compared to control, however, increase in the second experiment was extremely high (fold change = 17,480) compared to the first (fold change = 358). Of note, this order was presented by single member (*A. muciniphila*). Thus, we also analyzed the abundance of this microbe via specific PCR and confirmed differences between two experiments in vancomycin treated groups (0.04 and 9794.4 ng DNA *A. muciniphila*/ g cecal content in the first and second experiments, respectively).

Cocktail, ampicillin, and vancomycin treated samples showed similar patterns of change at the OTU level as compared to the microbiome of control samples (**Figure 4**), which may be related to the fact that only these antibiotic treatments were able to change GTT-AUC (**Figure 1**). *Prevotella* sp. (OTU\_189721) was the most abundant OTU in control (median abundance across two groups (24%), neomycin (38%), and metronidazole (17.5%)



**FIGURE 4 |** Heatmap of differentially abundant OTUs in the different groups. The represented OTUs passed the same statistical criteria as that used for phenotypes. The rows (OTUs) were clustered based on the abundance in the first experiment and the same row order was used for the second experiment. The rows were scaled, and orange and blue colors indicate increased and decreased abundance, respectively. The ids and taxa names for the rows (OTUs) are provided in **Supplementary Table S3**. The boxes indicate OTU clusters with the same taxonomy: Boxes 1 (113/117), 2 (43/46), and 3 (43/45) primarily contained Clostridiales (phylum: Firmicutes), while Box 4 (38/38) had Enterobacteriales (phylum: Proteobacteria).

treated samples. Enterobacteriaceae family (OTU\_1111294) was the most abundant in cocktail (38%) and vancomycin (28%), and the third most abundant in ampicillin (14%) treated samples. *Bacteroides uniformis* (OTU\_589071) was the most abundant upon ampicillin treatment (22.8%), while *A. muciniphila* was the second most abundant in vancomycin treated samples (17.5%) (**Supplementary Table S4**).

## Microbes Are Associated with Changed Phenotypes

Gut microbiota can control the expression of many genes in the small intestine (Larsson et al., 2012). Therefore, we asked whether the antibiotic-induced changes in the microbiome were potentially connected to the observed changes in gene expression. We constructed a transkingdom network using all groups, consisting of genes, metabolic parameters, and OTUs, to identify candidate interactions whereby microbes can mediate changes in systemic glucose tolerance and found 131 OTU-phenotype edges.

To focus on microbe-phenotype relationships that are not affected by type of antibiotics, we retained the 40 edges (**Figure 5**) that maintained the same sign of correlation coefficient between the various groups of both experiments and consistent with potential causal relations (Dong et al., 2015; Morgun et al., 2015; Greer R.L. et al., 2016; Rodrigues et al., 2017) in at least one group of both experiments. Overall, this means that while a strength of OTU-phenotype interaction may be weak for a particular antibiotic group, this interaction may still be important in mediating effects of antibiotics on the host in general. The

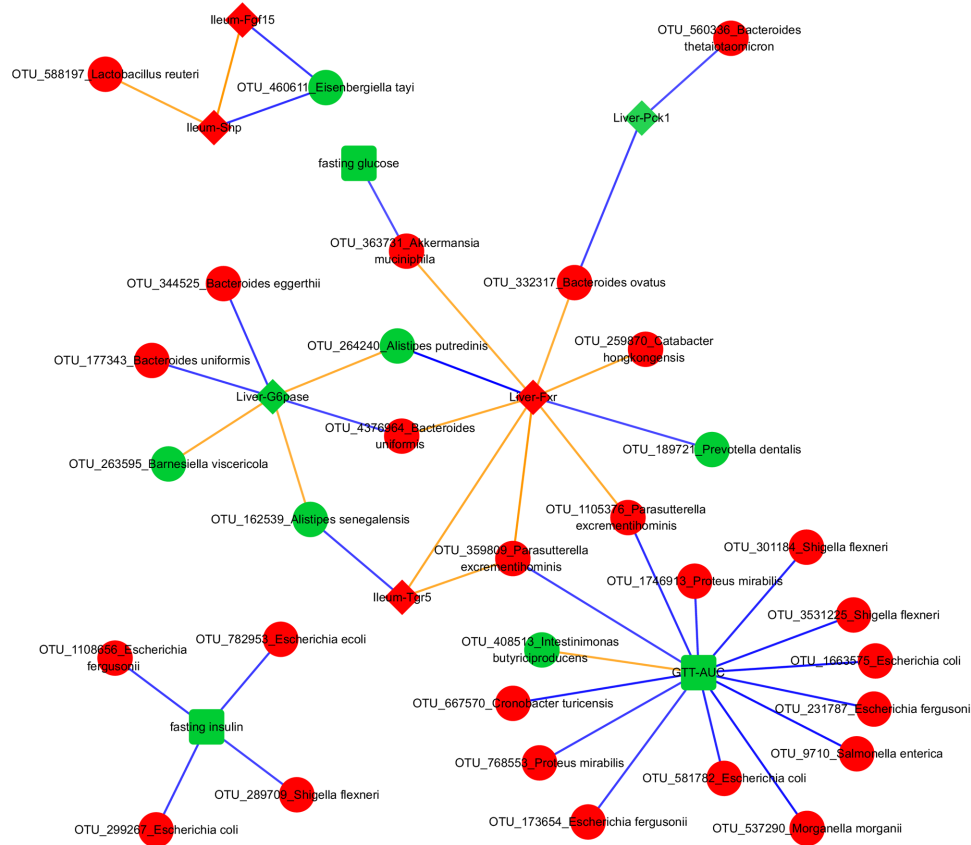
abundance of a microbe and its strength of correlation with a phenotype are expected to be crucial in mediating the effects, hence these 40 edges were ranked using a score that takes into account the maximum per-group OTU abundance and the median per-group correlation strength with a phenotype (**Figure 6**; See formula in section “Materials and Methods”).

NCBI BLAST on an OTU's Greengenes reference sequence was used to obtain its (closest) species level identification. Interestingly, associations between *A. muciniphila* with fasting glucose and liver Fxr showed as the top interactions suggesting a possible mechanism through which this bacterium can mediate systemic changes in glucose metabolism. *Proteus mirabilis* was negatively correlated with GTT-AUC. *Bacteroides uniformis* was positively and negatively correlated with hepatic Fxr and G6pase, respectively. Importance of phenotypes in the network was also determined by degrees of connectedness (degree) and BC. GTT-AUC (degree = 14, BC score = 0.65) and liver Fxr (degree = 8, BC score = 0.60) were the highly connected metabolic parameter and gene, respectively, as well as the key nodes in the largest connected component of the network. Overall, it suggests that gut microbiota potentially influences the liver metabolic genes and systemic metabolic parameters and mediates the effects of antibiotics on host phenotypes.

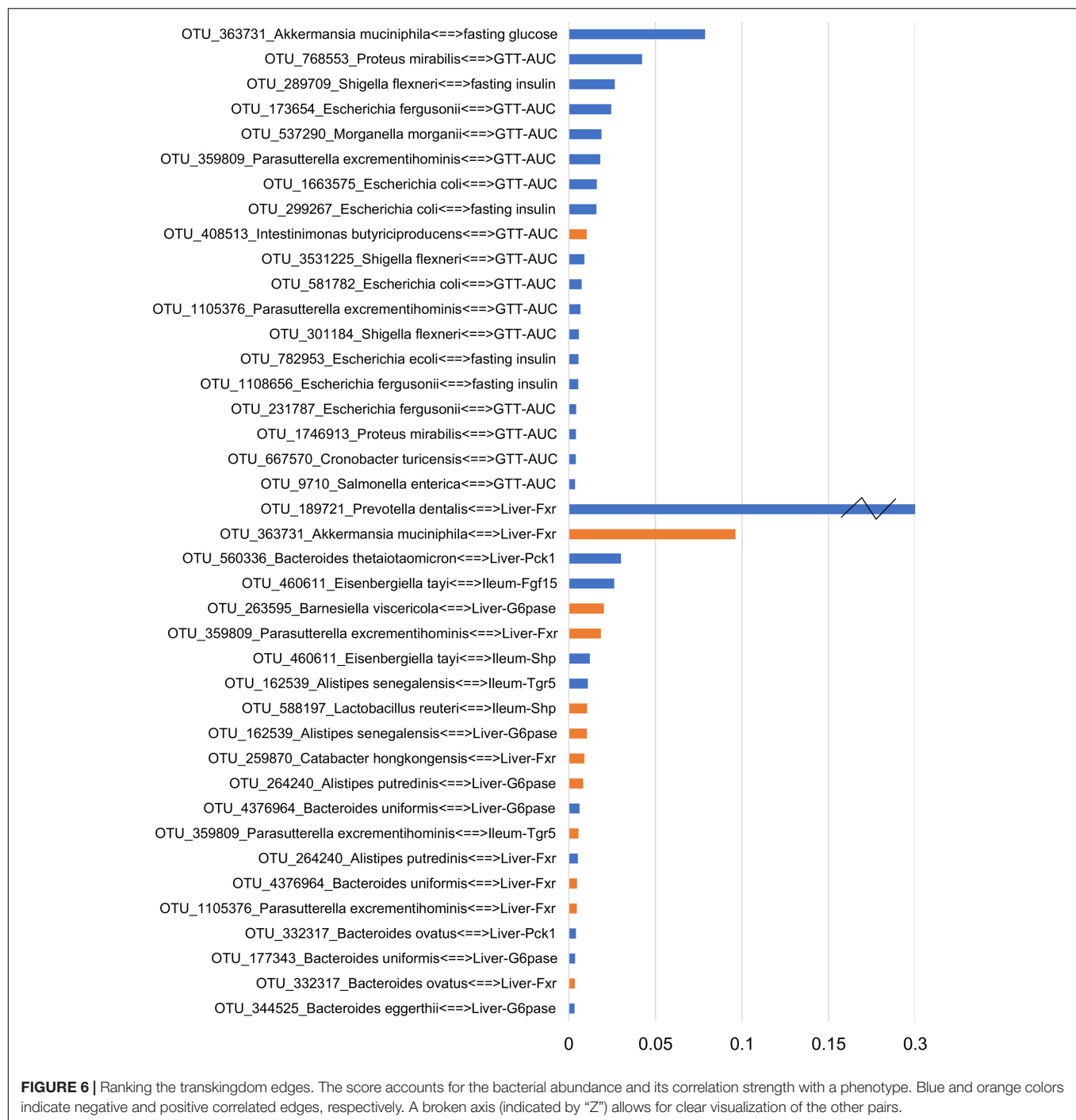
## DISCUSSION

Germ-free Swiss Webster mice showed improved glucose metabolism (Caesar et al., 2012), suggesting that microbiota





The expression of key genes from the glucose and bile acid metabolism pathways were measured, since bile acid signaling



plays an important role in glucose homeostasis (Nguyen and Bouscarel, 2008; Trauner et al., 2010; Prawitt et al., 2011; Chiang, 2013; Nie et al., 2015; Trabelsi et al., 2016). We observed well-known (and therefore expected) relationships between the tissue-specific expression patterns of different genes themselves and with the systemic metabolic parameters in vancomycin and ampicillin. Low hepatic Fxr causes increased gluconeogenesis (Ma et al., 2006) and bile acid synthesis (Duran-Sandoval et al., 2004), while increased liver Fxr (Li and Guo, 2015) and intestinal

Fgf15 (Holt et al., 2003) suppress hepatic bile acid synthesis (Kong et al., 2012) and regulate hepatic glucose metabolism (Potthoff et al., 2011). Also, increased liver Fxr represses G6pase (Yamagata et al., 2004; Ma et al., 2006; Zhang et al., 2006), Pck1 (De Fabiani et al., 2003; Yamagata et al., 2004; Ma et al., 2006), and like repressed ileum Fxr (Jiang et al., 2015b), improves glucose tolerance (Ma et al., 2006; Zhang et al., 2013), similar to our results. Along the same lines, mice treated with Fgf15 showed improved glucose metabolism (Zhou et al., 2017) and

increased intestinal Fgf15 expression represses liver G6pase and Pck1, key enzymes for liver gluconeogenesis (Potthoff et al., 2011). In line with our observations in vancomycin treatment, Fxr agonist obeticholic acid (OCA; Intercept Pharmaceuticals, New York, NY, United States) increased mRNA levels of Fgf15 and Tgr5 in the ileum of C57BL/6J mice without increase in ileum Fxr (Pathak et al., 2017), and an increase in ileum Fgf15 expression decreased plasma glucose levels even with low insulin levels (Potthoff et al., 2011). Supporting our results from ampicillin treatment, a study showed that treating mice on high fat diet with antibiotics cocktail inhibited Fxr signaling in the ileum but not in the liver, and observed decreased expression of Shp in the ileum (Jiang et al., 2015a). Furthermore, Fxr and Shp mRNA in the ileum were also reduced in germ-free Swiss Webster mice on chow diet compared to the conventionally raised (untreated) group (Sayin et al., 2013). Overall, these studies along with ours support the idea that bile acids repress gluconeogenesis (Modica et al., 2010). However, it is interesting to see that while metronidazole, neomycin, and cocktail do not show the above changes in gene expression there is still improvement in fasting glucose and/or glucose tolerance, suggesting that microbiota might play an even bigger role in mediating the effects of antibiotics on phenotypes through additional mechanisms not explored here.

In fact, microbiota can change the expression of many genes in the ileum (Larsson et al., 2012). Their study observed down-regulated Hk2, G6pase, and Shp in the ileum of germ-free mice when compared to conventionally raised mice on chow diet, supporting our results from cocktail (ampicillin, or vancomycin) treatment. Also, some of the changes in gut microbiota to antibiotics that we observe in our data, e.g., the increased Verrucomicrobiales following vancomycin (Hansen et al., 2012) and Enterobacteriales increase upon treatment with ampicillin or vancomycin (Ubeda et al., 2010), are well documented.

While the effects of microbes on systemic glucose tolerance in lean subjects are rarely studied, their ability to influence glucose metabolism is well-recognized (De Vadder et al., 2016). A good example of well-established causal relations between specific bacteria and glucose metabolism is beneficial effect of *A. muciniphila*. For example, it was shown that *A. muciniphila* was able to delay the onset of diabetes in the vancomycin treated mice (Hansen et al., 2012). Furthermore, multiple studies demonstrated that this bacterium can improve glucose metabolism in animal models and in humans (Everard et al., 2013; Zhang et al., 2013; Joyce and Gahan, 2014; Shin et al., 2014; Anhe et al., 2015; Dao et al., 2016; Greer R.L. et al., 2016). It might not be surprising that the negative correlation between *A. muciniphila* and glucose levels was detected as one of the top ranked edges in our unbiased transkingdom network, thus, providing extra confidence for our results about less investigated bacteria inferred in our analyses.

Our predictions provide insights into host-microbial interactions. For instance, our result of *Bacteroides uniformis* being correlated with hepatic G6pase and Fxr might indicate potential mechanisms by which this bacterium improves

glucose tolerance (Gauffin Cano et al., 2012). Also, it was shown that colonization with *Bacteroides thetaiotaomicron* makes mice leaner comparing controls despite similar levels of food consumption. The ability and preference of *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* to utilize polysaccharide rich diet (McNulty et al., 2013) may explain these effects. However, our result of the negative correlation between abundance of these two bacteria and Pck1 (simple sugar forming gluconeogenic enzyme) may suggest the effect of these bacteria on liver gluconeogenesis. Similarly, *P. mirabilis* is predicted to have a negative interaction with GTT-AUC in our study, but shown to be positively correlated in rats with and without high fat diet (Lecomte et al., 2015). This disagreement may be explained by different physiological pathways dominating in the same bacterial species in different host that has been clearly shown for other bacteria (Oh et al., 2010). Overall, our study offers testable hypothesis regarding critical microbe-phenotype associations.

## CONCLUSION

We show that antibiotics alter systemic glucose metabolism in lean mice. In addition to reporting changes in the microbiota, expression of key genes from the glucose and bile acid metabolism pathways, and concomitant systemic metabolic measures, we delineate potential mechanisms by which microbes mediate these effects. While there is a general understanding of the different players and mechanisms of microbiome-mediated regulation of the glycemic response (Tremaroli and Backhed, 2012; Devaraj et al., 2013; Cani et al., 2014; Hartstra et al., 2015; Janssen and Kersten, 2015; Parekh et al., 2015; Sanz et al., 2015; Boulange et al., 2016; Marchesi et al., 2016; Stenman et al., 2016; Suez et al., 2016), a lot remains to be understood, especially in terms of identifying the precise pathways operating in host-microbiome interactions. Overall, our data strongly suggests that antibiotics affect systemic glucose metabolism via shaping gut microbial communities and consequently regulating gene expression programs in intestine and liver. Yet, treatment of germfree mice with antibiotics as well as colonization of germ-free mice with antibiotic modified microbiota are required to fully support above statement. Also, it is doubtful that different antibiotics use the same mechanisms of gene expression and microbiota changes to affect systemic glucose tolerance, and the limited number of samples per group makes it difficult to obtain antibiotic-specific mechanisms. Furthermore, while the taxonomical assignments of 16S rRNA sequencing of current study present natural challenges, further studies employing shotgun metagenomics sequencing will allow to overcome this limitation. Finally, our experimental design followed by a data-driven, systems biology approach of network analysis offers consistent and statistically significant interactions that may be integral in mediating the host-microbiome communication. Furthermore, this approach is a useful hypothesis generating strategy and future experimentation can help investigate the distinct mechanisms in the different antibiotics and eventually lead to personalized medicine (Zmora et al., 2016).

## AUTHOR CONTRIBUTIONS

AM and NS conceived the original idea, designed and supervised the experiments, analyses, and writing. RG conceived the original idea, designed and performed the experiments, and supervised the writing. RR designed and performed the analyses, and drafted the manuscript. XD performed the analyses. JW, MG, and KD performed the experiments and analysis, respectively. All authors wrote the manuscript, read and approved the final draft submitted.

## ACKNOWLEDGMENTS

This research was supported by startup funds for AM and NS from Oregon State University (OSU), United States; NIH U01 AI109695 (AM), deLaubenfels Comparative Health Research and Education Fund (NS), and R01 DK103761 (NS).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02306/full#supplementary-material>

**FIGURE S1** | Dot plots with mean and error bars showing (A) body weight or (B–E) expression of the phenotypes across the different groups. The red and blue

colors indicate experiments one and two, respectively. These phenotypes did not show statistically significant differences upon antibiotics treatment compared to untreated control mice.

**FIGURE S2** | Boxplots showing the Shannon diversity index. Asterisk indicate statistically significant differences upon antibiotics treatment compared to untreated control mice: same direction of (abx/control) fold change in both experiments, individual  $p$ -value < 2% in each experiment, Fisher's combined  $p$ -value < 0.1% and FDR < 0.1%.

**FIGURE S3** | PCoA plot showing weighted UniFrac distance for cecal microbiota in the control and antibiotics treated mice. Each circle indicates a sample.

**FIGURE S4** | Taxonomic plots showing bacterial abundance across the different groups at the order level.

**FIGURE S5** | Taxonomic plots showing bacterial abundance across the different samples at the phylum level.

**FIGURE S6** | Taxonomic plots showing bacterial abundance across the different samples at the order level.

**TABLE S1** | Primers for the genes tested in this study.

**TABLE S2** | Alpha diversity metrics on the unrarefied and rarefied OTU tables in the two experiments.

**TABLE S3** | Levels of Firmicutes and Bacteroidetes in the two experiments. Asterisks indicate parameters that show statistically significant differences upon antibiotics treatment compared to untreated control mice: same direction of (abx/control) fold change in both experiments, individual  $p$ -value < 20% in each experiment, Fisher's combined  $p$ -value < 5% and FDR < 10%.

**TABLE S4** | The ID, Greengenes taxonomy, and the median frequency per group for the OTUs in the same order as of the heatmap.

## REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Anhe, F. F., Roy, D., Pilon, G., Dudonne, S., Matamoros, S., Varin, T. V., et al. (2015). A polyphenol-rich cranberry extract protects from diet-induced obesity, insulin resistance and intestinal inflammation in association with increased *Akkermansia* spp. population in the gut microbiota of mice. *Gut* 64, 872–883. doi: 10.1136/gutjnl-2014-307142
- Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284. doi: 10.1093/bioinformatics/btm554
- Boulange, C. L., Neves, A. L., Chilloux, J., Nicholson, J. K., and Dumas, M. E. (2016). Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med.* 8:42. doi: 10.1186/s13073-016-0303-2
- Caesar, R., Reigstad, C. S., Backhed, H. K., Reinhardt, C., Ketonen, M., Lunden, G. O., et al. (2012). Gut-derived lipopolysaccharide augments adipose macrophage accumulation but is not essential for impaired glucose or insulin tolerance in mice. *Gut* 61, 1701–1707. doi: 10.1136/gutjnl-2011-301689
- Cani, P. D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A. M., Delzenne, N. M., et al. (2008). Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes Metab. Res. Rev.* 57, 1470–1481. doi: 10.2337/db07-1403
- Cani, P. D., Geurts, L., Matamoros, S., Plovier, H., and Duparc, T. (2014). Glucose metabolism: focus on gut microbiota, the endocannabinoid system and beyond. *Diabetes Metab.* 40, 246–257. doi: 10.1016/j.diabet.2014.02.004
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.1303
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8
- Carvalho, B. M., Guadagnini, D., Tsukumo, D. M. L., Schenka, A. A., Latuf-Filho, P., Vassallo, J., et al. (2012). Modulation of gut microbiota by antibiotics improves insulin signalling in high-fat fed mice. *Diabetologia* 55, 2823–2834. doi: 10.1007/s00125-012-2648-4
- Chakraborti, C. K. (2015). New-found link between microbiota and obesity. *World J. Gastrointest. Pathophysiol.* 6, 110–119. doi: 10.4291/wjgp.v6.i4.110
- Chiang, J. Y. L. (2013). “Bile acid metabolism and signaling,” in *Comprehensive Physiology*, ed. D. M. Pollock (Hoboken, NJ: John Wiley & Sons, Inc.).
- Dao, M. C., Everard, A., Aron-Wisniewsky, J., Sokolowska, N., Prifti, E., Verger, E. O., et al. (2016). *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut* 65, 426–436. doi: 10.1136/gutjnl-2014-308778
- De Fabiani, E., Mitro, N., Gilardi, F., Caruso, D., Galli, G., and Crestani, M. (2003). Coordinated control of cholesterol catabolism to bile acids and of gluconeogenesis via a novel mechanism of transcription regulation linked to the fasted-to-fed cycle. *J. Biol. Chem.* 278, 39124–39132. doi: 10.1074/jbc.M305079200
- De La Cochetiere, M. F., Durand, T., Lepage, P., Bourreille, A., Galmiche, J. P., and Dore, J. (2005). Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. *J. Clin. Microbiol.* 43, 5588–5592. doi: 10.1128/JCM.43.11.5588-5592.2005
- De Vadder, F., Kovatcheva-Datchary, P., Zitoun, C., Duchamp, A., Backhed, F., and Mithieux, G. (2016). Microbiota-produced succinate improves glucose homeostasis via intestinal gluconeogenesis. *Cell Metab.* 24, 151–157. doi: 10.1016/j.cmet.2016.06.013
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05



- Dethlefsen, L., Huse, S., Sogin, M. L., and Relman, D. A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLOS Biol.* 6:e280. doi: 10.1371/journal.pbio.0060280
- Dethlefsen, L., and Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4554–4561. doi: 10.1073/pnas.1000087107
- Devaraj, S., Hemarajata, P., and Versalovic, J. (2013). The human gut microbiome and body metabolism: implications for obesity and diabetes. *Clin. Chem.* 59, 617–628. doi: 10.1373/clinchem.2012.187617
- Dong, X., Yambartsev, A., Ramsey, S. A., Thomas, L. D., Shulzhenko, N., and Morgun, A. (2015). Reverse enGENEering of regulatory networks from big data: a roadmap for biologists. *Bioinform. Biol. Insights* 9, 61–74. doi: 10.4137/BBI.S12467
- Duran-Sandoval, D., Mautino, G., Martin, G., Percevault, F., Barbier, O., Fruchart, J. C., et al. (2004). Glucose regulates the expression of the farnesoid X receptor in liver. *Diabetes* 53, 890–898. doi: 10.2337/diabetes.53.4.890
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J. P., Druart, C., Bindels, L. B., et al. (2013). Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9066–9071. doi: 10.1073/pnas.1219451110
- Fisher, R. A. (1932). “Statistical methods for research workers,” in *Biological Monographs and Manuals*, 5 Edn, eds F. A. E. Crew and D. W. Cutler (Edinburgh: Oliver and Boyd).
- Francino, M. P. (2015). Antibiotics and the human gut microbiome: dysbioses and accumulation of resistances. *Front. Microbiol.* 6:1543. doi: 10.3389/fmicb.2015.01543
- Fujisaka, S., Ussar, S., Clish, C., Devkota, S., Dreyfuss, J. M., Sakaguchi, M., et al. (2016). Antibiotic effects on gut microbiota and metabolism are host dependent. *J. Clin. Invest.* 126, 4430–4443. doi: 10.1172/JCI86674
- Gauffin Cano, P., Santacruz, A., Moya, A., and Sanz, Y. (2012). Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. *PLOS ONE* 7:e41079. doi: 10.1371/journal.pone.0041079
- Gower, J. (1998). “Principal coordinate analysis,” in *Encyclopedia of Biostatistics*, eds T. Coulton and P. Armitage (Hoboken, NJ: John Wiley and Sons Inc.), 3514–3518.
- Greer, R., Dong, X., Morgun, A., and Shulzhenko, N. (2016). Investigating a holobiont: microbiota perturbations and transkingdom networks. *Gut Microbes* 7, 126–135. doi: 10.1080/19490976.2015.1128625
- Greer, R. L., Dong, X., Moraes, A. C., Zielke, R. A., Fernandes, G. R., Peremyslova, E., et al. (2016). *Akkermansia muciniphila* mediates negative effects of IFN $\gamma$  on glucose metabolism. *Nat. Commun.* 7:13329. doi: 10.1038/ncomms13329
- Greer, R. L., Morgun, A., and Shulzhenko, N. (2013). Bridging immunity and lipid metabolism by gut microbiota. *J. Allergy Clin. Immunol.* 132, 253–263. doi: 10.1016/j.jaci.2013.06.025
- Hansen, C. H., Krych, L., Nielsen, D. S., Vogensen, F. K., Hansen, L. H., Sorensen, S. J., et al. (2012). Early life treatment with vancomycin propagates *Akkermansia muciniphila* and reduces diabetes incidence in the NOD mouse. *Diabetologia* 55, 2285–2294. doi: 10.1007/s00125-012-2564-7
- Hartstra, A. V., Bouter, K. E., Backhed, F., and Nieuwdorp, M. (2015). Insights into the role of the microbiome in obesity and type 2 diabetes. *Diabetes Care* 38, 159–165. doi: 10.2337/dc14-0769
- Holt, J. A., Luo, G., Billin, A. N., Bisi, J., McNeill, Y. Y., Kozarsky, K. F., et al. (2003). Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. *Genes Dev.* 17, 1581–1591. doi: 10.1101/gad.1083503
- Horikoshi, M., and Tang, Y. (2016). *ggfortify: Data Visualization Tools for Statistical Analysis Results*. Available at: <https://cran.r-project.org/web/packages/ggfortify/index.html>
- Human Microbiome Project Consortium (2012). A framework for human microbiome research. *Nature* 486, 215–221. doi: 10.1038/nature11209
- Hwang, I., Park, Y. J., Kim, Y. R., Kim, Y. N., Ka, S., Lee, H. Y., et al. (2015). Alteration of gut microbiota by vancomycin and bacitracin improves insulin resistance via glucagon-like peptide 1 in diet-induced obesity. *FASEB J.* 29, 2397–2411. doi: 10.1096/fj.14-265983
- Jakobsson, H. E., Jernberg, C., Andersson, A. F., Sjolund-Karlsson, M., Jansson, J. K., and Engstrand, L. (2010). Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLOS ONE* 5:e9836. doi: 10.1371/journal.pone.0009836
- Janssen, A. W., and Kersten, S. (2015). The role of the gut microbiota in metabolic health. *FASEB J.* 29, 3111–3123. doi: 10.1096/fj.14-269514
- Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J. K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J.* 1, 56–66. doi: 10.1038/ismej.2007.3
- Jiang, C., Xie, C., Li, F., Zhang, L., Nichols, R. G., Krausz, K. W., et al. (2015a). Intestinal farnesoid X receptor signaling promotes nonalcoholic fatty liver disease. *J. Clin. Invest.* 125, 386–402. doi: 10.1172/JCI76738
- Jiang, C., Xie, C., Lv, Y., Li, J., Krausz, K. W., Shi, J., et al. (2015b). Intestine-selective farnesoid X receptor inhibition improves obesity-related metabolic dysfunction. *Nat. Commun.* 6:10166. doi: 10.1038/ncomms10166
- Joyce, S. A., and Gahan, C. G. (2014). The gut microbiota and the metabolic health of the host. *Curr. Opin. Gastroenterol.* 30, 120–127. doi: 10.1097/MOG.000000000000039
- Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C. J., Fagerberg, B., et al. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498, 99–103. doi: 10.1038/nature12198
- Kasai, C., Sugimoto, K., Moritani, I., Tanaka, J., Oya, Y., Inoue, H., et al. (2015). Comparison of the gut microbiota composition between obese and non-obese individuals in a Japanese population, as analyzed by terminal restriction fragment length polymorphism and next-generation sequencing. *BMC Gastroenterol.* 15:100. doi: 10.1186/s12876-015-0330-2
- Komsta, L. (2011). *Outliers: Tests for Outliers*. Available at: <http://www.r-project.org>
- Kong, B., Wang, L., Chiang, J. Y., Zhang, Y., Klaassen, C. D., and Guo, G. L. (2012). Mechanism of tissue-specific farnesoid X receptor in suppressing the expression of genes in bile-acid synthesis in mice. *Hepatology* 56, 1034–1043. doi: 10.1002/hep.25740
- Larsson, E., Tremaroli, V., Lee, Y. S., Koren, O., Nookaew, I., Fricker, A., et al. (2012). Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut* 61, 1124–1131. doi: 10.1136/gutjnl-2011-301104
- Lecomte, V., Kaakoush, N. O., Maloney, C. A., Raipuria, M., Huinao, K. D., Mitchell, H. M., et al. (2015). Changes in gut microbiota in rats fed a high fat diet correlate with obesity-associated metabolic parameters. *PLOS ONE* 10:e0126931. doi: 10.1371/journal.pone.0126931
- Li, G., and Guo, G. L. (2015). Farnesoid X receptor, the bile acid sensing nuclear receptor, in liver regeneration. *Acta Pharm. Sin. B* 5, 93–98. doi: 10.1016/j.apsb.2015.01.005
- Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005
- Ma, K., Saha, P. K., Chan, L., and Moore, D. D. (2006). Farnesoid X receptor is essential for normal glucose homeostasis. *J. Clin. Invest.* 116, 1102–1109. doi: 10.1172/JCI25604
- Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D., Hirschfield, G. M., Hold, G., et al. (2016). The gut microbiota and host health: a new clinical frontier. *Gut* 65, 330–339. doi: 10.1136/gutjnl-2015-309990
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139
- McNulty, N. P., Wu, M., Erickson, A. R., Pan, C., Erickson, B. K., Martens, E. C., et al. (2013). Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. *PLOS Biol.* 11:e1001637. doi: 10.1371/journal.pbio.1001637
- Membrez, M., Blancher, F., Jaquet, M., Bibiloni, R., Cani, P. D., Burcelin, R. G., et al. (2008). Gut microbiota modulation with norfloxacin and ampicillin enhances glucose tolerance in mice. *FASEB J.* 22, 2416–2426. doi: 10.1096/fj.07-102723

- Mikkelsen, K. H., Allin, K. H., and Knop, F. K. (2016). Effect of antibiotics on gut microbiota, glucose metabolism and body weight regulation: a review of the literature. *Diabetes. Obes. Metab.* 18, 444–453. doi: 10.1111/dom.12637
- Mikkelsen, K. H., Frost, M., Bahl, M. I., Licht, T. R., Jensen, U. S., Rosenberg, J., et al. (2015). Effect of antibiotics on gut microbiota, gut hormones and glucose metabolism. *PLOS ONE* 10:e0142352. doi: 10.1371/journal.pone.0142352
- Modica, S., Gadaleta, R. M., and Moschetta, A. (2010). Deciphering the nuclear bile acid receptor FXR paradigm. *Nucl. Recept. Signal.* 8:e005. doi: 10.1621/nrs.08005
- Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., and Schaffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics* 24, 1757–1764. doi: 10.1093/bioinformatics/btn322
- Morgun, A., Dzutsev, A., Dong, X., Greer, R. L., Sexton, D. J., Ravel, J., et al. (2015). Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut* 64, 1732–1743. doi: 10.1136/gutjnl-2014-308820
- Murri, M., Leiva, I., Gomez-Zumaquero, J. M., Tinahones, F. J., Cardona, F., Soriguer, F., et al. (2013). Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC Med.* 11:46. doi: 10.1186/1741-7015-11-46
- Nguyen, A., and Bouscarel, B. (2008). Bile acids and signal transduction: role in glucose homeostasis. *Cell. Signal.* 20, 2180–2197. doi: 10.1016/j.cellsig.2008.06.014
- Nie, Y. F., Hu, J., and Yan, X. H. (2015). Cross-talk between bile acids and intestinal microbiota in host metabolism and health. *J. Zhejiang Univ. Sci. B* 16, 436–446. doi: 10.1631/jzus.B1400327
- Oh, P. L., Benson, A. K., Peterson, D. A., Patil, P. B., Moriyama, E. N., Roos, S., et al. (2010). Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *ISME J.* 4, 377–387. doi: 10.1038/ismej.2009.123
- Pallav, K., Dowd, S. E., Villafuerte, J., Yang, X., Kabbani, T., Hansen, J., et al. (2014). Effects of polysaccharopeptide from *Trametes versicolor* and amoxicillin on the gut microbiome of healthy volunteers: a randomized clinical trial. *Gut Microbes* 5, 458–467. doi: 10.4161/gmic.29558
- Panda, S., El khader, I., Casellas, F., Lopez Vivancos, J., Garcia Cors, M., Santiago, A., et al. (2014). Short-term effect of antibiotics on human gut microbiota. *PLOS ONE* 9:e95476. doi: 10.1371/journal.pone.0095476
- Pang, J., Rhodes, D. H., Pini, M., Akasheh, R. T., Castellanos, K. J., Cabay, R. J., et al. (2013). Increased adiposity, dysregulated glucose metabolism and systemic inflammation in Galectin-3 KO mice. *PLOS ONE* 8:e57915. doi: 10.1371/journal.pone.0057915
- Parekh, P. J., Balart, L. A., and Johnson, D. A. (2015). The influence of the gut microbiome on obesity, metabolic syndrome and gastrointestinal disease. *Clin. Transl. Gastroenterol.* 6:e91. doi: 10.1038/ctg.2015.16
- Pathak, P., Liu, H., Boehme, S., Xie, C., Krausz, K. W., Gonzalez, F., et al. (2017). Farnesoid X receptor induces Takeda G-protein receptor 5 cross-talk to regulate bile acid synthesis and hepatic metabolism. *J. Biol. Chem.* 292, 11055–11069. doi: 10.1074/jbc.M117.784322
- Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* 10, 1200–1202. doi: 10.1038/nmeth.2658
- Perez-Cobas, A. E., Gosalbes, M. J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K., et al. (2013). Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* 62, 1591–1601. doi: 10.1136/gutjnl-2012-303184
- Potthoff, M. J., Boney-Montoya, J., Choi, M., He, T., Sunny, N. E., Satapati, S., et al. (2011). FGF15/19 regulates hepatic glucose metabolism by inhibiting the CREB-PGC-1 $\alpha$  pathway. *Cell Metab.* 13, 729–738. doi: 10.1016/j.cmet.2011.03.019
- Prawitt, J., Caron, S., and Staels, B. (2011). Bile acid metabolism and the pathogenesis of type 2 diabetes. *Curr. Diab. Rep.* 11, 160–166. doi: 10.1007/s11892-011-0187-x
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi: 10.1038/nature11450
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* 118, 229–241. doi: 10.1016/j.cell.2004.07.002
- Raymond, F., Ouameur, A. A., Deraspe, M., Iqbal, N., Gingras, H., Dridi, B., et al. (2016). The initial state of the human gut microbiome determines its reshaping by antibiotics. *ISME J.* 10, 707–720. doi: 10.1038/ismej.2015.148
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y. F., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rodrigues, R. R., Shulzhenko, N., and Morgun, A. (2017). Transkingdom networks: a systems biology approach to identify causal members of host-microbiota interactions. (accepted). Available at: <https://arxiv.org/pdf/1709.05701.pdf>
- Saeidi, N., Meoli, L., Nestoridi, E., Gupta, N. K., Kvas, S., Kucharczyk, J., et al. (2013). Reprogramming of intestinal glucose metabolism and glycemic control in rats after gastric bypass. *Science* 341, 406–410. doi: 10.1126/science.1235103
- Sanz, Y., Olivares, M., Moya-Perez, A., and Agostoni, C. (2015). Understanding the role of gut microbiome in metabolic disease risk. *Pediatr. Res.* 77, 236–244. doi: 10.1038/pr.2014.170
- Sayin, S. I., Wahlstrom, A., Felin, J., Jantti, S., Marschall, H. U., Bamberg, K., et al. (2013). Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* 17, 225–235. doi: 10.1016/j.cmet.2013.01.003
- Schneeberger, M., Everard, A., Gomez-Valades, A. G., Matamoros, S., Ramirez, S., Delzenne, N. M., et al. (2015). *Akkermansia muciniphila* inversely correlates with the onset of inflammation, altered adipose tissue metabolism and metabolic disorders during obesity in mice. *Sci. Rep.* 5:16643. doi: 10.1038/srep16643
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shin, N. R., Lee, J. C., Lee, H. Y., Kim, M. S., Whon, T. W., Lee, M. S., et al. (2014). An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut* 63, 727–735. doi: 10.1136/gutjnl-2012-303839
- Stenman, L. K., Burcelin, R., and Lahtinen, S. (2016). Establishing a causal link between gut microbes, body weight gain and glucose metabolism in humans - towards treatment with probiotics. *Benef. Microbes* 7, 11–22. doi: 10.3920/BM2015.0069
- Suez, J., Shapiro, H., and Elinav, E. (2016). Role of the microbiome in the normal and aberrant glycemic response. *Clin. Nutr. Exp.* 6, 59–73. doi: 10.1016/j.jclynex.2016.01.001
- Thomas, C., Pellicciari, R., Pruzanski, M., Auwerx, J., and Schoonjans, K. (2008). Targeting bile-acid signalling for metabolic diseases. *Nat. Rev. Drug Discov.* 7, 678–693. doi: 10.1038/nrd2619
- Trabelsi, M. S., Lestavel, S., Staels, B., and Collet, X. (2016). Intestinal bile acid receptors are key regulators of glucose homeostasis. *Proc. Nutr. Soc.* 76, 192–202. doi: 10.1017/S0029665116002834
- Trauner, M., Claudel, T., Fickert, P., Moustafa, T., and Wagner, M. (2010). Bile acids as regulators of hepatic lipid and glucose metabolism. *Dig. Dis.* 28, 220–224. doi: 10.1159/000282091
- Tremaroli, V., and Backhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature* 489, 242–249. doi: 10.1038/nature11552
- Ubeda, C., Taur, Y., Jenq, R. R., Equinda, M. J., Son, T., Samstein, M., et al. (2010). Vancomycin-resistant *Enterococcus* domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* 120, 4332–4341. doi: 10.1172/JCI43918
- Utzschneider, K. M., Kratz, M., Damman, C. J., and Hullarg, M. (2016). Mechanisms linking the gut microbiome and glucose metabolism. *J. Clin. Endocrinol. Metab.* 101, 1445–1454. doi: 10.1210/jc.2015-4251
- Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* 2:16. doi: 10.1186/2047-217X-2-16
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., et al. (2016). *gplots: Various R Programming Tools for Plotting Data*. Available at: <https://CRAN.R-project.org/package=gplots>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-0-387-98141-3

- Wu, H., Tremaroli, V., and Backhed, F. (2015). Linking microbiota to human diseases: a systems biology perspective. *Trends Endocrinol. Metab.* 26, 758–770. doi: 10.1016/j.tem.2015.09.011
- Yamagata, K., Daitoku, H., Shimamoto, Y., Matsuzaki, H., Hirota, K., Ishida, J., et al. (2004). Bile acids regulate gluconeogenic gene expression via small heterodimer partner-mediated repression of hepatocyte nuclear factor 4 and Foxo1. *J. Biol. Chem.* 279, 23158–23165. doi: 10.1074/jbc.M314322200
- Yambartsev, A., Perlin, M. A., Kovchegov, Y., Shulzhenko, N., Mine, K. L., Dong, X., et al. (2016). Unexpected links reflect the noise in networks. *Biol. Direct* 11:52. doi: 10.1186/s13062-016-0155-0
- Zhang, X., Shen, D., Fang, Z., Jie, Z., Qiu, X., Zhang, C., et al. (2013). Human gut microbiota changes reveal the progression of glucose intolerance. *PLOS ONE* 8:e71108. doi: 10.1371/journal.pone.0071108
- Zhang, Y., Lee, F. Y., Barrera, G., Lee, H., Vales, C., Gonzalez, F. J., et al. (2006). Activation of the nuclear receptor FXR improves hyperglycemia and hyperlipidemia in diabetic mice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1006–1011. doi: 10.1073/pnas.0506982103
- Zhou, M., Luo, J., Chen, M., Yang, H., Learned, R. M., DePaoli, A. M., et al. (2017). Mouse species-specific control of hepatocarcinogenesis and metabolism by FGF19/FGF15. *J. Hepatol.* 66, 1182–1192. doi: 10.1016/j.jhep.2017.01.027
- Zmora, N., Zeevi, D., Korem, T., Segal, E., and Elinav, E. (2016). Taking it personally: personalized utilization of the human microbiome in health and disease. *Cell Host Microbe* 19, 12–20. doi: 10.1016/j.chom.2015.12.016

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Rodrigues, Greer, Dong, DSouza, Gurung, Wu, Morgun and Shulzhenko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Intriguing Evolutionary Journey of Enteroinvasive *E. coli* (EIEC) toward Pathogenicity

Martina Pasqua<sup>1</sup>, Valeria Michelacci<sup>2</sup>, Maria Letizia Di Martino<sup>1†</sup>, Rosangela Tozzoli<sup>2</sup>, Milena Grossi<sup>1</sup>, Bianca Colonna<sup>1</sup>, Stefano Morabito<sup>2</sup> and Gianni Prosseda<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

David A. Rasko,  
University of Maryland, Baltimore,  
United States  
Alessandra Polissi,  
Università degli Studi di Milano, Italy  
Antonio Juárez,  
University of Barcelona, Spain

### \*Correspondence:

Gianni Prosseda  
gianni.prosseda@uniroma1.it

### † Present address:

Maria Letizia Di Martino,  
Science for Life Laboratory,  
Department of Medical Biochemistry  
and Microbiology, Uppsala University,  
Uppsala, Sweden

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 18 September 2017

**Accepted:** 20 November 2017

**Published:** 05 December 2017

### Citation:

Pasqua M, Michelacci V,  
Di Martino ML, Tozzoli R, Grossi M,  
Colonna B, Morabito S and  
Prosseda G (2017) The Intriguing  
Evolutionary Journey of Enteroinvasive  
*E. coli* (EIEC) toward Pathogenicity.  
Front. Microbiol. 8:2390.  
doi: 10.3389/fmicb.2017.02390

<sup>1</sup> Istituto Pasteur Italia, Department of Biology and Biotechnology "C. Darwin", Sapienza Università di Roma, Rome, Italy,

<sup>2</sup> European Union Reference Laboratory for *Escherichia coli*, Department of Veterinary Public Health and Food Safety, Istituto Superiore di Sanità, Rome, Italy

Among the intestinal pathogenic *Escherichia coli*, enteroinvasive *E. coli* (EIEC) are a group of intracellular pathogens able to enter epithelial cells of colon, multiply within them, and move between adjacent cells with a mechanism similar to *Shigella*, the etiological agent of bacillary dysentery. Despite EIEC belong to the same pathotype of *Shigella*, they neither have the full set of traits that define *Shigella* nor have undergone the extensive gene decay observed in *Shigella*. Molecular analysis confirms that EIEC are widely distributed among *E. coli* phylogenetic groups and correspond to bioserotypes found in many *E. coli* serogroups. Like *Shigella*, also in EIEC the critical event toward a pathogenic life-style consisted in the acquisition by horizontal gene transfer of a large F-type plasmid (pINV) containing the genes required for invasion, intracellular survival, and spreading through the intestinal mucosa. In *Shigella*, the ample gain in virulence determinants has been counteracted by a substantial loss of functions that, although important for the survival in the environment, are redundant or deleterious for the life inside the host. The pathoadaptation process that has led *Shigella* to modify its metabolic profile and increase its pathogenic potential is still in infancy in EIEC, although maintenance of some features typical of *E. coli* might favor their emerging relevance as intestinal pathogens worldwide, as documented by recent outbreaks in industrialized countries. In this review, we will discuss the evolution of EIEC toward *Shigella*-like invasive forms going through the epidemiology, including the emergence of new virulent strains, their genome organization, and the complex interactions they establish with the host.

**Keywords:** pathogenic *E. coli*, enteroinvasive *E. coli* (EIEC), *Shigella*, bacterial evolution, emerging EIEC

**Abbreviations:** DAMP, damage-associated molecular pattern; DEC, diarrheagenic *E. coli*; HGT, horizontal gene transfer; H-NS, heat-stable nucleoid-structuring protein; IL, interleukin; IS, insertion sequence; MAPK, mitogen-activated protein kinase; NLR, Nod-like receptor; PAI, pathogenicity island; PAMP, pathogen-associated molecular pattern; pINV, virulence plasmid; PMNL, polymorphonuclear leukocytes; PRR, pattern recognition receptor; SHI, *Shigella* pathogenicity island; SRL, *Shigella* resistance locus; sRNA, small RNA; T2SS, type II secretion system; T3SS, type III secretion system; TLR, Toll-like receptor; TNF, tumor necrosis factor.



## INTRODUCTION

*Escherichia coli* is not only a harmless commensal of the human and animal intestine but also a major cause of morbidity and mortality (Kaper et al., 2004; Wirth et al., 2006). Indeed, many pathogenic *E. coli* have been described as cause of diseases both in healthy and immunocompromised individuals. Based on the specific virulence factors and pathogenicity processes, pathogenic *E. coli* have been subdivided into different pathogroups, that can be broadly grouped as DEC (or intestinal) or extraintestinal *E. coli* (ExPEC) (Kaper et al., 2004; Croxen and Finlay, 2009; Gomes et al., 2016). DEC include at least six major pathotypes differing in virulence mechanisms, infectious processes, and damages provoked to the target cells: enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC), as well as adherent invasive *E. coli* (AIEC), a recently identified pathotype. As for ExPEC, the most common strains belong to two different pathotypes targeting different body compartments: uropathogenic *E. coli* (UTI) and neonatal meningitis *E. coli* (NMEC).

The presence of so many different pathotypes exemplifies the remarkable plasticity of *E. coli* genome, which is characterized by an extremely large pangenome of approximately 20,000 genes in contrast to a common core of about 1700 genes (Rasko et al., 2008; Touchon et al., 2009). Those that vary among different pathogenic *E. coli* strains have been acquired by intense HGT and are often conveyed by mobile genetic elements (Touchon et al., 2009; Dobrindt et al., 2010; van Elsas et al., 2011).

Among the DEC pathotypes, EIEC are etiological agents of bacillary dysentery in humans, particularly in low-income countries (Croxen et al., 2013; Gomes et al., 2016). The pathogenesis of EIEC infection is characterized by the ability of bacteria to invade the human colonic mucosa, conferred by the expression of chromosomal and plasmid-borne genes (Harris et al., 1982; Sansonetti et al., 1982; Hale et al., 1983; Kaper et al., 2004). Following penetration into colonic epithelial cells, EIEC replicate intracellularly and spread to adjacent cells causing the inflammatory destruction of the intestinal epithelial barrier. This provokes the characteristic dysentery syndrome, usually self-limiting, characterized by the presence of blood, mucus, and leukocytes in stools (DuPont et al., 1971; O'Brien et al., 1979; Taylor et al., 1988). The clinical illness caused by EIEC is similar to that induced by *Shigella* spp. (Formal and Hornick, 1978; Small and Falkow, 1988), with whom they are closely related in their virulence and other phenotypic properties (Kopecko et al., 1985; Lan et al., 2004). Notwithstanding the similarities in the invasion mechanisms, the infectious dose of EIEC has been observed to be much higher than that of *Shigella* and the diseases caused by EIEC appear in some cases to be milder (DuPont et al., 1971).

Despite several studies, whether EIEC are precursors of the “full blown” pathogen *Shigella*, or not is still under debate. In this review, we will attempt at tracing the evolutionary pathway of EIEC considering their epidemiology, the complex mechanisms of their interaction with host cells, the key steps that could have characterized their evolution from a commensal life style toward

pathogenicity, and the organization of their genome, including the description of the major traits of emerging EIEC clones.

## EPIDEMIOLOGY OF ENTEROINVASIVE *E. coli* (EIEC)

The first report of an EIEC strain dates back at 1947 (Ewing and Gravatti, 1947). At that time, it was defined as “paracolon bacillus” but the strain was later identified as an O124 *E. coli*. In the 1950s and 1960s, other *E. coli* strains, isolated from dysentery and initially classified as *Shigella manolovi*, *S. sofia*, *Shigella* strain 13, and *S. metadysenteriae*, due to their ability to cause experimental keratoconjunctivitis in guinea pigs, were later renamed as EIEC (Manolov, 1959; Rowe et al., 1977; Edwards and Ewing, 1986). Their biochemical characters were first described in 1967 (Sakazaki et al., 1967; Trabulsi et al., 1967).

Enteroinvasive *E. coli* and *Shigella* spp. share several phenotypic and genotypic characteristics, often making the discrimination between the two genera challenging (Silva et al., 1980; Toledo and Trabulsi, 1983; Bando et al., 1998; Lan and Reeves, 2002; Pavlovic et al., 2011; van den Beld and Reubsæet, 2012), especially in case of shared serogroups. This difficulty biases the interpretation of the epidemiological information available, hindering the evaluation of the real burden of EIEC infections. As a matter of fact, both EIEC and *Shigella* spend much of their life cycle within the eukaryotic cells, possessing the ability to use nutrients coming from the host environment. Similarly to *Shigella*, most EIEC strains are unable to decarboxylate lysine, lack the ability to ferment lactose, and are generally non-motile, with the exception of strains belonging to a few serogroups (Silva et al., 1980; Farmer et al., 1985; Bando et al., 1998; Casalino et al., 2003; Tozzoli and Scheutz, 2014).

A limited set of serotypes have been assigned to EIEC, namely O28ac:H-, O29:H-, O112ac:H-, O115:H-, O121:H-, O124:H-, O124:H7, O124:H30, O124:H32, O135:H-, O136:H-, O143:H-, O144:H-, O144:H25, O152:H-, O159:H-, O159:H2, O164:H-, O167:H-, O167:H4, O167:H5, O173:H-, and recently O96:H19 (Voeroes et al., 1964; Silva et al., 1980; Gomes et al., 1987, 2016; Orskov et al., 1991; Matsushita et al., 1993; Escher et al., 2014; Tozzoli and Scheutz, 2014; Michelacci et al., 2016; Newitt et al., 2016). Some of these EIEC-associated O antigens, such as O28, O112ac, O121, O124, O143, O144, O152, and O167, are identical to O antigens present in *Shigella* spp. (Cheasty and Rowe, 1983; Tozzoli and Scheutz, 2014).

Enteroinvasive *E. coli*-infected humans seem to be the major source of infection, as no animal reservoirs have been identified, and transmission uses mainly the oral–fecal route. Although EIEC infections occur worldwide, these are particularly common in low-income countries where poor general hygiene favors their spreading (Chatterjee and Sanyal, 1984; Beutin et al., 1997; Kaper et al., 2004; Vieira et al., 2007).

Enteroinvasive *E. coli* incidence has been estimated in several countries, and it differs depending on the region (Gomes et al., 2016). Discrepancies among some of the reports can be observed, probably due to the difficulty in discriminating between *Shigella* and EIEC. In certain countries of Latin America and Asia,

namely Chile, Thailand, India, and Brazil, EIEC were found to be common diarrheagenic pathogens (Chatterjee and Sanyal, 1984; Faundez et al., 1988; Echeverria et al., 1992; Blake et al., 1993; Levine et al., 1993), with frequent reports of asymptomatic infected subjects excreting the pathogen (Beutin et al., 1997). In industrialized countries, EIEC infections have been mainly described as travel-related, being reported in returning travelers from high-incidence countries (Wanger et al., 1988; Beutin et al., 1997; Svenungsson et al., 2000). Occasionally, food and water sources have been identified as vehicles of infection, but usually as a secondary contamination by a human source (Tozzoli and Scheutz, 2014).

Enteroinvasive *E. coli* cause sporadic cases of infection but have been implicated in outbreaks as well, sometimes involving large numbers of cases. In the 1970s a huge outbreak, affecting 387 patients and linked to cheese contaminated with an O124 *E. coli* strain, occurred in United States (Marier et al., 1973). Recently, an increase of cases of infections linked to an emerging EIEC clone has been observed in Europe, where in 2012 a large and severe outbreak of bloody diarrhea in Italy involving more than 100 individuals was reported (Escher et al., 2014; Pettengill et al., 2015). An EIEC O96:H19 strain, a serotype never described before for EIEC, was isolated and the suspected source of infection was traced to cooked vegetables (Escher et al., 2014). During the outbreak investigation an EIEC O96:H19 strain was also isolated from two asymptomatic food handlers working in the canteen linked with the outbreak, supporting the hypothesis of a secondary contamination of the vegetables during post-cooking handling procedures (Escher et al., 2014). In 2014, two linked outbreaks of gastrointestinal disease occurred in the United Kingdom, involving more than 100 cases of infection. One of the episodes was associated to the consumption of contaminated salad vegetables and, again, an O96:H19 EIEC was isolated from some of the patients and from vegetable samples (Newitt et al., 2016). Finally, an EIEC belonging to the same serotype was isolated in a case of traveler's diarrhea in Spain in 2013 (Michelacci et al., 2016). Pheno-genotypic characterization of the strains involved in the three episodes suggests that the EIEC O96:H19 could be emerged as a result of the recent acquisition of the invasion plasmid by an *E. coli* strain (Michelacci et al., 2016).

## THE INVASIVE PROCESS

Similarly to *Shigella*, EIEC are responsible of bacillary dysentery (Taylor et al., 1988). However, the disease caused by EIEC is usually less severe than that induced by *Shigella* (DuPont et al., 1971). Following the discovery that EIEC strains carry a pINV plasmid identical to that of *Shigella* (Harris et al., 1982; Sansonetti et al., 1982; Hale et al., 1983) and that they can display a *Shigella*-like invasive behavior (Hale et al., 1985; Small and Falkow, 1988; Taylor et al., 1988), *in vitro* and *in vivo* studies have been extensively focused on *Shigella*, providing in-depth knowledge about its pathogenicity/virulence mechanisms. In recent years, the pathogenicity of EIEC has gained new interest and comparative analyses between EIEC and *Shigella* have been

performed, aimed at understanding the different clinical outcome severity of the two infections (Moreno et al., 2009; Bando et al., 2010; Sanchez-Villamil et al., 2016). Here we first present the invasive process as it has been inferred from studies on *S. flexneri*. Then, we address what it is known about the difference between these two enteroinvasive bacteria.

In order to gain access to intestinal epithelia, bacteria first transit from the lumen to the submucosa by preferentially entering M cells in Peyer's patches. After endocytosis by M cells bacteria are transcytosed toward the M cell pocket, where they meet, and are phagocytosed by resident macrophages. *Shigella* infection of macrophages is accompanied by the release of T3SS effectors and components that are recognized as PAMPs by NLRs, ultimately leading to pyroptosis with the release of proinflammatory cytokines, IL-1 $\beta$  and IL-18 (Ashida et al., 2011). The induction of macrophage cell death is pivotal for bacteria to invade enterocytes, though pyroptosis is a form of cell death that induces a massive inflammatory response. Once released from dying macrophages, invasive bacteria infect the neighboring enterocytes by entering through the basolateral surface. Here they are enclosed into a vacuole that is rapidly disrupted freeing them into the cytosol. Subsequently, the bacteria multiply and, using actin-based motility, spread to adjacent cells (Schroeder and Hilbi, 2008).

Inside epithelial cells, bacterial PAMPs and DAMPs are detected by various PRRs, including TLRs and NLRs, which stimulate host defense signal pathways such as those involving MAPKs and NF- $\kappa$ B leading to the secretion of proinflammatory cytokines (e.g., IL-8 and TNF- $\alpha$ ) (Takeuchi and Akira, 2010). These molecules induce the recruitment of phagocytic cells to the infection site, initially facilitating the invasion process and eventually clearing the bacterial pathogens. In order to maximize invasion and permanence and save the replicative niche in epithelial cells, invading *Shigella* modulate host cell responses throughout the infection process by secreted effectors (Killackey et al., 2016). Induction of a very early inflammatory response upon invasion of epithelial cells is functional to bacterial spreading as it results in recruitment of polymorphonuclear leucocytes (PMNL), which migrate across the epithelium destabilizing the intercellular junctions and increasing the surface available for bacterial entry into target cells (Ashida et al., 2011). Several T3SS effectors, such as OspB, OspC1, and OspZ (Zurawski et al., 2009; Ambrosi et al., 2015; Mattock and Blocker, 2017), contribute to promote inflammation at early stages of the infection process. They mainly act by enhancing activation of MAPK and NF- $\kappa$ B pathways, which are involved in the control of the production of PMNL chemoattractants, including IL-8, whose secretion triggers PMNL migration in a basolateral to apical direction causing epithelial barrier disruption. However, though this early inflammatory response is essential to initiate infection, it would also contribute toward rapidly clearing the infecting agents. Thus, to establish infection, at later stages *Shigella* must overcome the host innate response. This is achieved by delivering T3SS effectors, whose function is aimed mainly at inhibiting MAPK and NF- $\kappa$ B signaling pathways with the consequent decrease of inflammatory chemokine and cytokine production (Killackey et al., 2016; Mattock and Blocker, 2017).

An important obstacle *Shigella* must tackle during the invasion of the epithelial tissue is host cell targeting and degradation by autophagy. Several studies have demonstrated that *Shigella* are particularly exposed to autophagy targeting only when they are associated to cell membranes. Two bacterial factors, IcsB and VirA, have been implicated in bacterial evasion of autophagy targeting by interfering with LC3 recruitment and by allowing bacteria to escape from LC3-positive vacuoles (Ogawa et al., 2005; Baxt and Goldberg, 2014; Campbell-Valois et al., 2015).

Typically, intracellular pathogens need to save their host to establish a successful infection. As part of their pathogenic mechanism *Shigella* employ several countermeasures to avoid premature cell death to maintain their epithelial replicative niche. The early stage of infection is characterized by induction of DNA damage and genotoxic stress, which lead to activation of p53 and stimulation of apoptosis. Apoptotic cell death is prevented by the activity of the T3SS effectors VirA and IpgD, which promote p53 degradation and activate the PI3K/Akt pro-survival pathway, respectively, and by the pilus component protein FimA, which inhibits cytochrome c release by mitochondria (Mattock and Blocker, 2017).

As discussed above, EIEC share many aspects of the *Shigella* infection process that involves crossing of intestinal epithelial barrier, killing of resident macrophage cells, invasion of enterocytes, intra-cellular replication, and dissemination from cell to cell without extracellular steps (Croxen and Finlay, 2009). Moreover, EIEC express the same virulence factors found in *Shigella* (Parsot, 2005). However, the infectious dose required for EIEC to cause disease is higher than that of *Shigella* and the disease caused by EIEC appears to be milder (DuPont et al., 1971), suggesting differences between EIEC and *Shigella* in sensing and shaping the host environment, which, in turn, would influence the pathways toward virulence. To date only few studies have investigated the differences in the infectiveness between EIEC and *Shigella*. Moreno et al. (2009) detailed for the first time the relationship between the expression of some genes crucial for the infection process and the reduced ability of EIEC to cause disease. This is well supported by their Serény tests in guinea pigs, showing how the signs of keratoconjunctivitis induced by *Shigella* appear earlier and are more severe as compared to those caused by EIEC. Using an epithelial cell model, the authors also demonstrate that, although *Shigella* and EIEC display similar invasion ability, EIEC disseminates less efficiently, producing smaller plaques in plaque assays. As compared to *Shigella* the overall behavior of EIEC apparently reflects a reduced expression of key virulence genes, during both invasion and cell-to-cell spreading, except for *virF* that is expressed at higher levels by intracellular EIEC than *Shigella* during the dissemination step. This apparent discrepancy may be explained in the light of recent results showing that *Shigella virF* is transcribed into two mRNAs, with the shortest one encoding a smaller protein that negatively regulates transcription of full-length mRNA and, consequently, the expression of the VirF regulator (Di Martino et al., 2016b). Since in the real-time PCR experiments carried out by Moreno et al. (2009) *virF* expression was assayed by using primers that did not discriminate

between the two mRNAs, comparative *virF* expression studies between *Shigella* and EIEC deserve further investigations to deeper analyze potential differences.

A more recent work has compared the host cell response to infection by different *E. coli* pathotypes, including EIEC, and by *Shigella*. The kinetic of NF- $\kappa$ B and ERK1/2 activation in HT-29 epithelial cells shows only a slightly higher p65 phosphorylation after 4 h of infection with *Shigella* as compared with EIEC. Conversely, although following a similar kinetics, the accumulation of phosphorylated ERK1/2 is much higher in cells infected with EIEC at 4 h post-infection. Despite these differences, HT-29 cells infected with EIEC or *Shigella* release comparable amounts of cytokines, as IL-8 and TNF- $\alpha$  with similar kinetics (Sanchez-Villamil et al., 2016). The phosphorylation of ERK1/2 and p38 is controlled by the phosphothreonine lyase activity of OspF, to which both anti-inflammatory (Arbibe et al., 2007) and pro-inflammatory roles (Reiterer et al., 2011) have been attributed. Since both *Shigella* and EIEC express OspF, it is reasonable that additional factors are involved in determining the different ERK1/2 phosphorylation profile and the outcome of MAPK activation.

The key step in invasion of the epithelial cells resides in the ability of EIEC and *Shigella* to escape from macrophages after phagocytosis by induction of caspase 1-dependent cell death. It has been reported that, as compared to *Shigella*, EIEC have a decreased capacity to escape from murine J774 macrophages and are less efficient in cell killing during the first 4 h of infection (Bando et al., 2010). This likely depends on differences in the expression of some virulence genes. In particular, as compared to *Shigella* the expression of the *ipaC* gene is reduced in intracellular EIEC at all the time points after infection. As for the release of pro- and anti-inflammatory cytokines (as TNF- $\alpha$ , IL-1, and IL-10) by infected cells, contrasting results exist. While no significant differences between EIEC and *Shigella*-infected J774 cells (Bando et al., 2010) have been reported, other studies carried out using human THP-1 cells differentiated into macrophages (Sanchez-Villamil et al., 2016) have shown that *Shigella* infection results in higher secretion of both pro-inflammatory and anti-inflammatory cytokines.

To date, banking on the modest amount of data available from *in vitro* infection of macrophage-like cells and epithelial cells, the milder disease caused by EIEC appears to be mainly associated to a lower expression of key virulence genes involved in phagosomal escape inside host cells and in dissemination among epithelial cells (Moreno et al., 2009; Bando et al., 2010). There are no obvious differences in the inflammatory response by epithelial cells, at least as far as the secretion of IL-8 and TNF- $\alpha$  is concerned, neither at early nor at late times of infection. Despite this cytokine profile, the activation state of ERK1/2 MAPK seems to be more elevated in epithelial cells infected with EIEC than in those infected with *Shigella* (Sanchez-Villamil et al., 2016). Deeper investigations will clarify to what extent this may depend on differences in manipulating certain cell signaling pathways and on differences in the activity of bacterial factors involved therein.



## THE MAJOR VIRULENCE TRAIT OF EIEC: THE LARGE VIRULENCE PLASMID pINV

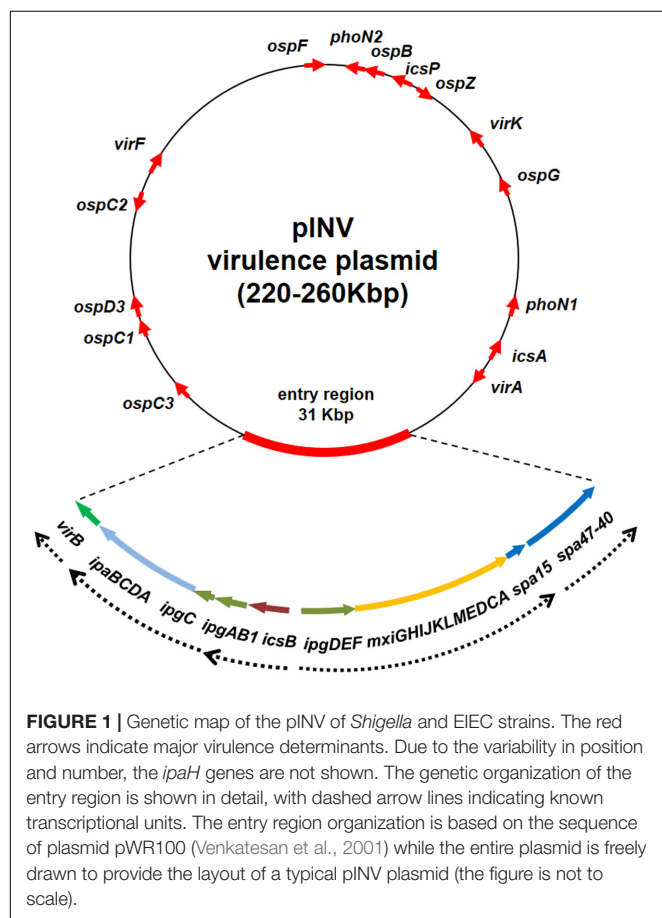
The evolution of *E. coli* toward pathogenic phenotypes has been determined, as in many other bacterial pathogens, mainly by two mechanisms: the acquisition of virulence genes by HGT as parts of plasmids, phages, transposons, or PAI and the loss or modification of genes of the core genome. While the first mechanism plays a crucial role in the colonization of a new host environment, the latter, known as pathoadaptation, strongly contributes to drive the evolution of bacteria toward a more pathogenic phenotype (Kaper et al., 2004; Dobrindt et al., 2010).

It is widely acknowledged that, as in *Shigella*, in EIEC the critical event in the transition toward a pathogenic lifestyle has been the acquisition of a large F-type plasmid (pINV) which encodes the molecular machinery required for invasion, survival, and diffusion of the bacterium within the host (Harris et al., 1982; Sansonetti et al., 1982; Hale et al., 1983; Small and Falkow, 1988; **Figure 1**). The pINV plasmid has been found only in the *Shigella*/EIEC pathotype and its loss is a very rare event, which determines an avirulent phenotype.

The genetic organization of the pINV is very complex (Johnson and Nolan, 2009). As a matter of fact these plasmids are

made up of a mosaic of genes of various origins and harbor traces of four different plasmids (Buchrieser et al., 2000; Venkatesan et al., 2001; Escobar-Páramo et al., 2003). pINV isolated from EIEC share wide regions of high structural and functional homology and are interchangeable with those isolated from *Shigella* strains (Hale et al., 1983; Lan et al., 2001; Johnson and Nolan, 2009). pINV share with IncFIIA plasmids high homology in the regions involved in replication (*rep*) and conjugation (*tra*) (Makino et al., 1988) and stable inheritance of pINV is ensured by the presence of several plasmid segregation and maintenance systems (Lan et al., 2001). Due to large deletions in the *tra* region, pINV are not capable of self-transfer by conjugation, but they can be mobilized by other conjugative plasmids. All over the plasmid genome, an astonishing number of ISs is present as a mixture of complete and incomplete IS elements repeated several times, confirming the relevant role played by ISs in pINV assembly and evolution (Buchrieser et al., 2000; Venkatesan et al., 2001). Most ISs are related to known elements while others represent novel ISs. Among the latter, ISEc11, an IS belonging to the IS1111 family, is widespread and functional in pINV from EIEC while only defective copies are present in the *Shigella* pINV plasmids (Prosseda et al., 2006).

In the pINV there is only one large (31 kb) region, which does not host any IS elements. This is the so-called entry region, which displays a PAI-like structure (Buchrieser et al., 2000; Venkatesan et al., 2001). It is composed by two large, divergently transcribed gene clusters coding for a T3SS apparatus (Mxi and Spa), for most of its effectors (IpaB, IpaC, and IpaD) with their chaperons (IpgA, IpgC, IpgE, and Spa15), and for two transcriptional regulators (VirB and MxiE), both required for the activation of most virulence genes (Schroeder and Hilbi, 2008; **Figure 1**). The entry region is extremely conserved among *Shigella* and EIEC pINV plasmids (Lan et al., 2001). Albeit it had been initially proposed as a PAI, likely acquired in a single recombination event, it lacks the presence of flanking tRNA sequences and at least remnants of a recombinase-encoding gene. It is therefore unclear if the acquisition of the entry region has occurred independently from its insertion into tRNA sequences or if the absence of tRNA genes may have resulted from rearrangement following gene transfer. The latter hypothesis is supported by the fact that the entry region is flanked by truncated IS elements, suggesting that rearrangements may have occurred after its acquisition *en bloc* by the plasmid (Buchrieser et al., 2000). The T3SS encoded by the entry region plays a critical role in the bacterial invasive process, since it delivers a large number of effectors involved in the reorganization of the host cell actin cytoskeleton and in the modulation of cell signaling pathways to evade the host immune response (Mattock and Blocker, 2017). With the exception of few proteins of the IpaH family, which are chromosomally encoded, all T3SS effectors are encoded by pINV genes located within or outside the entry region. Since the entry region is highly conserved, the phylogenetic analysis of three of its genes (*ipgD*, *mxiC*, and *mxiA*) has allowed differentiating pINV from *Shigella* spp. and EIEC into two forms, A and B, with the first one predominantly associated with EIEC strains (Lan et al., 2001, 2004).





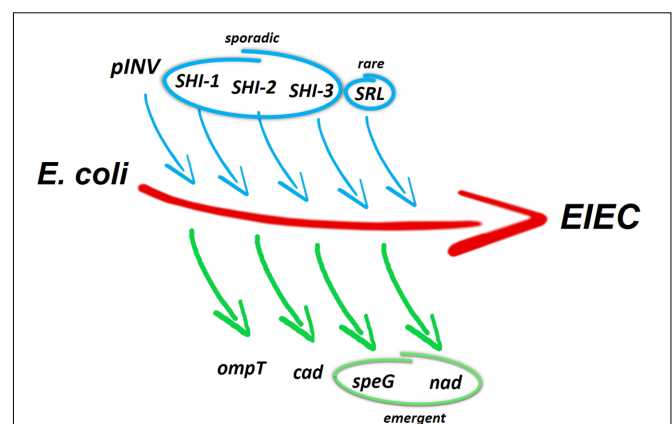
Besides the large PAI-like region, a small islet carries the genes coding for IcsA (a protein responsible for the bacterial motility inside the cytoplasm), VirA (a GTPase-activating protein), and RnaG (a regulatory sRNA negatively controlling *icsA* expression) (Giangrossi et al., 2010; Tran et al., 2011; Dong et al., 2012). Other genes encoding proteins crucial for the invasive process cover the pINV plasmid including the OspG and OspF proteins which interfere with the host innate immune response (Kim et al., 2005; Arbibe et al., 2007), the PhoN2 protein required for IcsA localization (Scribano et al., 2014), and the IpaH proteins which interfere with the host protein degradation (Ashida and Sasakawa, 2015; **Figure 1**). Moreover, in contrast to the other two virulence regulatory genes (*virB* and *mxiE*), the *virF* gene, coding for the primary virulence regulator, is located on a “desert island” surrounded by several IS sequences and far away from all other virulence genes, including those under its direct control, *virB* and *icsA* (Di Martino et al., 2016a). While the CG content of *virF* is only slightly lower as compared to that of the entry region (Buchrieser et al., 2000), its position suggests that it has been acquired independently to promote the expression of the virulence genes. VirF is also involved in the activation of some chromosomal genes, indicating that it acts as global regulator and that its acquisition by HGT has contributed to a reshaping of the core genome, easing the adaptation of bacteria to the host environment (Barbagallo et al., 2011; Leuzzi et al., 2015).

The mechanisms involved in the activation of the pINV virulence genes have been extensively studied both in EIEC and in *Shigella* (Dagberg and Uhlin, 1992; Prosseda et al., 2002). They rely on a sophisticated regulatory cascade involving global and specific regulators, encoded by both, pINV and the chromosome. Outside the human host, the nucleoid-associated protein H-NS represses each of the key promoters of the pINV virulence genes (Dorman, 2004). In response to environmental conditions found in the human intestine, the transcriptional activation of the invasive operons is triggered by an increased level of VirF counteracting H-NS repression at the *icsA* and *virB* promoters (Prosseda et al., 2004). Then VirB activates most operons within the entry region, including the gene for the last regulator (*mxiE*), as well as all other virulence genes scattered along the pINV genome, except *icsA*. Finally, MxiE, assisted by IpgC, activates the transcription of genes encoding the late effectors (Schroeder and Hilbi, 2008).

As in other pathogenic *E. coli*, also in EIEC the virulence genes are stably maintained on an extrachromosomal element (Johnson and Nolan, 2009). Nevertheless, it has been reported that the pINV of EIEC strain HN280 is able to integrate into the host chromosome and that integration results in silencing of all pINV-encoded virulence genes also under host temperature conditions (Zagaglia et al., 1991). Silencing was shown to depend on a severe reduction of *virB* transcription, likely dependent on the inability of VirF to counteract the negative control of H-NS at the *virB* promoter when it is chromosomally located (Colonna et al., 1995). This has led to the hypothesis that the presence of virulence genes on the pINV is the result of an evolutionary pathway toward the optimization of gene expression.

## EVOLUTION OF EIEC

The studies on the evolutionary origin of the *Shigella*/EIEC pathovar have led to two major hypotheses. The pINV could have been acquired only once by an ancestral *E. coli* that subsequently gave rise to the different *Shigella*/EIEC lineages (Escobar-Páramo et al., 2003; Zuo et al., 2013), as suggested by the inability of the plasmid to autonomously undergo horizontal transmission. Alternatively, the different *Shigella*/EIEC strains could have arisen from different *E. coli* that had acquired the pINV independently, e.g., from an unknown donor or from other *Shigella*/EIEC that already harbored it. This view is supported by the diversity of the genotypes within the *Shigella*/EIEC pathovar, revealed by phylogenetic analyses of chromosomal genes and by genome comparison (Pupo et al., 2000; Hazen et al., 2016; Pettengill et al., 2016). Besides the large pINV, several virulence genes have been acquired on the chromosome of *Shigella* and EIEC as part of PAIs (**Figure 2**). The PAIs described so far for *Shigella* (SHI islands) carry genes encoding different traits, including an enterotoxin and a cytotoxic protease (SHI-1) and systems involved in iron uptake and evasion of immune response (SHI-2 and SHI-3 in *S. boydii*), in the modification of O antigens (SHI-O) or in multi-drug resistance (SRL) (Schroeder and Hilbi, 2008). Recently, 20 genomes from EIEC belonging to different serotypes have been compared with those of reference strains belonging to diverse *E. coli* pathovars and *Shigella* species. This comparison highlights the existence of at least three distinct lineages containing only EIEC strains and suggests a convergent evolution of non-pathogenic *E. coli* toward invasive phenotype (Hazen et al., 2016). An *in silico* search for protein-encoding genes of SHI-1, SHI-2, SHI-3, SHI-O, and SRL indicates that, with



**FIGURE 2 |** Genetic events contributing to the evolution of EIEC from ancestral commensal *E. coli*. The acquisition of the pINV by HGT is a major evolutionary event toward pathogenicity. This can be accompanied by the sporadic acquisition of entire or incomplete SHI-1 PAI and incomplete SHI-2 and SHI-3 PAIs. Rarely, also incomplete SRL PAI are acquired by EIEC genomes. The absence of *ompT* and the loss of cadaverine synthesis (usually resulting from *cadC* silencing) counterbalance the gain of virulence-associated determinants. The inactivation of *speG* (involved in spermidine acetylation) and *nad* (involved in NAD biosynthesis) is regarded as emergent pathoadaptive mutations in EIEC.

the exception of SHI-O, portions of the other PAIs are present in EIEC genomes in a lineage-specific manner (Hazen et al., 2016). Interestingly, while a whole SHI-1 Island has never been detected in EIEC, SHI-1 fragments of different length have been found in all EIEC genomes. However, the ShET1 toxin genes, typically harbored by SHI-1 in *S. flexneri* genomes, were found only in EIEC strains of lineage 2. In the case of virulence genes associated with SHI-2, the *shiA* gene, involved in the reduction of the host inflammatory response, is absent in all EIEC lineages, while *shiD*, which provides immunity to colicins, is present in all EIEC of lineages 1 and 2. An entire SHI-3 PAI, typically associated with *S. boydii* strains, has been detected only in few EIEC strains of lineages 1 and 2, while portions of it, including the genes encoding for aerobactin-mediated iron uptake, are found in all three lineages. As for the large SRL PAI, widely distributed among *Shigella* spp. and containing a cluster of multiple antibiotic resistance determinants (Turner et al., 2003), only a few of its genes are present in EIEC genomes.

The variable presence of the PAIs in EIEC confirms the phylogenetic diversity among EIEC and *Shigella* and further supports the hypothesis that the EIEC pathovar has not a single origin but rather stems from multiple lineages (Hazen et al., 2016; Michelacci et al., 2016; Pettengill et al., 2016).

A significant complementary step toward the pathogenic lifestyle has been pathoadaptation, the inactivation, or loss of several chromosomal genes, which negatively interfere with the expression of virulence factors required for survival within the host. The antivirulence loci identified encode a broad spectrum of functions, confirming that adaptation to the new host environments is the result of long and ordered process targeting core genome determinants (Casalino et al., 2003; Di Martino et al., 2013b; Campilongo et al., 2014).

Despite the close similarity of the *Shigella* and EIEC pathogenicity process, it is well known that EIEC have a metabolic activity more similar to *E. coli* and have not undergone the intense gene decay observed in *Shigella* (Silva et al., 1980; Pettengill et al., 2016). It is therefore not surprising that the pathoadaptation in EIEC has not reached a level comparable to *Shigella* (Prosseda et al., 2012) and that most of the antivirulence loci characterized in *Shigella* are still encoding functional products in EIEC. One of the pathoadaptive mutations conserved both in EIEC and in *Shigella* is the deletion of the *ompT* gene, located within the defective lambdoid prophage DLP12 (Nakata et al., 1993; Figure 2). The OmpT protease triggers the degradation of IcsA protein and therefore negatively interferes with host cell invasion by drastically reducing the ability of *Shigella* to spread into adjacent epithelial cells. Considering that the loss of OmpT is widespread, it is as yet unclear if *E. coli* lineages that gave rise to the *Shigella*/EIEC pathovar have not hosted DLP12 *ab initio* or if the prophage has been excised during the pathoadaptation process (Bliven and Maurelli, 2012; Leuzzi et al., 2017).

Another typical pathoadaptive mutation of *Shigella* spp. is the inability to catabolize lysine, due to the silencing of lysine decarboxylase (LDC) activity (Prosseda et al., 2007). The LDC<sup>-</sup> phenotype, which is found also in most EIEC, is determined by mutations in the *cad* locus, which hamper the

synthesis of cadaverine. Cadaverine is a polyamine that interferes with pathogenicity by blocking the release of *Shigella* into the cytoplasm of the infected cells and inhibiting the migration of PMNL across the intestinal epithelium (Bliven and Maurelli, 2012). A detailed analysis of the molecular rearrangements occurred in the *cad* operon of several EIEC strains belonging to different serotypes (Casalino et al., 2003) has shown that, similarly to *Shigella*, the silencing of the *cad* locus has been accomplished through convergent evolution. In contrast to *Shigella*, in EIEC the *cad* region is colinear with the *E. coli* K12 chromosome and the lack of cadaverine synthesis is mainly due to the inactivation of the gene encoding the CadC transcriptional regulator (Casalino et al., 2010). By comparing the *cad* loci of EIEC and *Shigella*, it appears that the rearrangements occurred in EIEC are less severe compared to the complete erosion of the locus observed in several *Shigella* strains (Casalino et al., 2005; Prosseda et al., 2007; Figure 2). Indeed, despite the antivirulence role played by cadaverine (Fernandez et al., 2001), emerging O96:H19 EIEC strains still maintains an integer *cad* operon and exhibits a LDC<sup>+</sup> phenotype (Michelacci et al., 2016).

As compared to the commensal *E. coli* the polyamine profile of *Shigella* is affected not only by the lack of cadaverine but also by the marked accumulation of spermidine and by the loss of *N*-acetyl spermidine, the inert form of spermidine (Di Martino et al., 2013a). The increased spermidine content depends on the loss of the spermidine acetyltransferase (SAT), the enzyme encoded by the *speG* gene and responsible for the conversion of spermidine into *N*-acetylspermidine. In *Shigella* it has been demonstrated that a higher level of spermidine increases survival within macrophages and confers bacteria a higher resistance to oxidative stress (Barbagallo et al., 2011). Similarly to how observed for the *cad* locus, also *speG* silencing is the result of convergent evolution. A comparison of the polyamine profiles of several EIEC strains with those of *Shigella* and *E. coli* K12 has revealed that in EIEC major polyamines attain levels in-between those observed in *E. coli* and *Shigella*. Indeed, as compared to commensal *E. coli*, in EIEC intracellular putrescine is significantly increased and spermidine tends to be higher. Nevertheless, in contrast to *Shigella*, *N*-acetylspermidine is still present in most EIEC strains (Campilongo et al., 2014), indicating that the loss of *speG* is an emerging trait. However, when spermidine accumulation is induced in EIEC through deletion of the *speG* gene, survival within macrophages, as well as resistance to oxidative stress are increased (Campilongo et al., 2014). This confirms that the absence of SAT activity confers to intracellular bacteria like EIEC and *Shigella* an increased capability to defy antagonistic host environment. Moreover, the analysis of the polyamine profiles has revealed that the higher level of putrescine in EIEC is determined by increased transcription of *speC*, promoted by the lack of cadaverine. The *speC* gene encodes the enzyme converting L-ornithine into putrescine. On the basis of these observations it has been suggested (Campilongo et al., 2014) that during the transition toward the pathogenic phenotype, the modification of the polyamine profile might have been triggered by the loss of cadaverine, with the double effect of favoring the invasive process and increasing the putrescine level. Since putrescine is an important

intermediate in the synthesis of spermidine and, consequently, of *N*-acetylspermidine, its increase may in turn have caused higher levels of both polyamines. In this scenario the silencing of *speG*, which appears completed in *Shigella* but can be regarded as an ongoing process in EIEC, would represent the last step favoring further accumulation of spermidine and the disappearance of *N*-acetylspermidine.

Another noteworthy pathoadaptive mutation in *Shigella* is the requirement for exogenous nicotinic acid (NAD) due to inactivation of the *nad* genes (Prunier et al., 2007), required for *de novo* synthesis of NAD. Also in this case the inability to synthesize NAD is not a generalized feature among EIEC strains (Di Martino et al., 2013b). In those EIEC strains requiring NAD it has been shown that the preferential target in the pathoadaptation process is the *nadB* gene, inactivated through diverse strategies, involving point mutations or IS insertions.

Altogether, the picture emerging from the observations on pathoadaptive mutations suggests that EIEC might represent intermediates in the evolution toward a full-blown phenotype, with some mutational events still confined to *Shigella* (Figure 2). However, a recent whole-genome comparative analysis (Pettengill et al., 2016), performed on a large number of *Shigella* and EIEC genomes, indicates that *Shigella* and EIEC evolved independently. Nevertheless, the same authors proposed that, while EIEC as a group cannot be considered the ancestor to *Shigella*, some EIEC lineages might have been the *Shigella* ancestor.

## EMERGING ENTEROINVASIVE *Escherichia coli*

The recent outbreaks occurred in Europe caused by the EIEC O96:H19 led the scientific community to reconsider the role of EIEC infection in industrialized countries (Escher et al., 2014; Michelacci et al., 2016; Newitt et al., 2016). Such EIEC serotype had never been reported before 2012 and represents a new virulent emergent clone. The EIEC O96:H19 isolated from two outbreaks occurred in Italy and United Kingdom and from a sporadic case of disease reported in Spain were studied by whole genome sequencing (Pettengill et al., 2015; Michelacci et al., 2016). The genomic analysis confirmed that all the isolates belonged not only to the same unprecedented EIEC serotype, but also to the same sequence type (ST-99), never observed before in EIEC strains (Michelacci et al., 2016). The analysis of the distribution of virulence genes typical of EIEC and *Shigella* highlighted the presence in the three strains of the plasmid genes encoding the T3SS system and its effectors, as well as the master transcriptional regulators genes *virF* and *virB*. As for the chromosomally located virulence genes, the three isolates showed the presence of the genetic determinants of a T2SS and were all negative for those encoding the aerobactin system involved in iron uptake. Interestingly, none of the O96:H19 isolates was found to have undergone the process of pathoadaptation through accumulation of the mutations described in the literature for EIEC and *Shigella* (Bliven and Maurelli, 2012; Prosseda et al., 2012). Nevertheless, the three isolates were shown to display

minor differences. The plasmid profiles obtained through the genomic analysis highlighted the presence of five plasmids in the strains isolated in Spain and United Kingdom and three plasmids in that responsible of the Italian outbreak, with three plasmids in common in the three strains. Altogether, these observations strengthen the hypothesis of the emergence of a new virulent EIEC clone circulating in Europe.

Phenotypic analysis also highlighted peculiar properties of this EIEC clone, when compared to reference EIEC and *Shigella* strains. Biochemical characterization showed that the isolates displayed the LDC activity, confirming the lack of the related pathoadaptive mutations observed through genome analysis, and interestingly showed that the isolates retained the ability to ferment lactose (Michelacci et al., 2016), usually lacking in *Shigella* and in the majority of EIEC strains (Tozzoli and Scheutz, 2014). Generally, a better fitness was observed for the O96:H19 strains when comparing the growth curves with those of *Shigella* and reference EIEC strains (Michelacci et al., 2016). Moreover, swimming motility was observed for the strains from Italian and Spanish cases, which was instead completely absent in the strain from United Kingdom and in all the other EIEC and *Shigella* strains tested. Such phenotypic traits are not typical of intracellular pathogens such as EIEC and *Shigella*, while they are more common in *E. coli* strains, contributing to their great ability in surviving and adapting in different ecological niches.

These findings support the hypothesis of the evolution of EIEC and *Shigella* after the acquisition of the pINV by multiple lineages of commensal *E. coli*, followed by a multi-step adaptation process. Such an evolutionary pathway could be exemplified by EIEC ST-280 Clonal Complex, which could have been generated with the acquisition of the pINV plasmid by a commensal *E. coli* eventually evolving toward *Shigella* belonging to related clonal complexes (ST-149, 152, 243, 245, 250) (Wirth et al., 2006; Michelacci et al., 2016). The mechanism could have involved multiple events of pathoadaptive mutations, giving origin to the existing *Shigella* clones, specialized for intracellular survival with detriment of the ability to persist outside the host. A similar paradigm could also explain the emergence of other EIEC clones following the acquisition of pINV by other commensal *E. coli*. This event in some cases could be followed by the accumulation of pathoadaptive mutation, as it is the case of the EIEC strains belonging to ST-6 clonal complex, while some other clones could have maintained all the functions granting an efficient extracellular persistence, such as the EIEC O96:H19 belonging to ST-99 (Michelacci et al., 2016). The observed better fitness of EIEC O96:H19 in comparison with that of the other reference EIEC and *Shigella* strains could have favored its survival in the extracellular environment and allowed its overgrowth in the food vehicles, granting it a high potential as a foodborne pathogen, as demonstrated in the two large episodes occurred in Italy and United Kingdom (Michelacci et al., 2016; Newitt et al., 2016).

## CONCLUSION AND PERSPECTIVES

Genomics approaches in combination with phenotypic analyses have a strong potential toward the formulation of new intriguing



hypotheses on the ongoing evolution of EIEC. Currently available comparisons between EIEC and *Shigella* genomes support the need for a taxonomical revision moving the *Shigella* genus back within the *E. coli* species (Michelacci et al., 2016; Pettengill et al., 2016). As a matter of fact, *Shigella* clades are interspersed in clusters of *E. coli* genomes regardless of the bioinformatics approach used for the phylogenetic analysis (Sahl et al., 2015; Pettengill et al., 2016). In the light of recent studies, the organization of the EIEC genome appears to have been originated from multiple independent events (Hazen et al., 2016; Pettengill et al., 2016). This hypothesis finds even stronger evidence in the emergence of a novel EIEC clone belonging to O96:H19 serotype, which exhibits phenotypic traits more typical of *E. coli* than of reference EIEC or *Shigella* (Michelacci et al., 2016).

The acquisition of the plasmid may represent the first step in the emergence of new EIEC clones, but it is well known to be not sufficient for establishing the full pathogenicity (Sansonetti et al., 1983). In this context, it is of great

interest to deeper investigate on the role and relevance of functions that *Shigella* has lost in its route toward an intracellular life-style but that are still retained by most EIEC strains.

## AUTHOR CONTRIBUTIONS

BC, VM, SM, and GP proposed the idea of the review; MP, VM, MG, and RT wrote the review draft; MP, MM, and GP design the figures; BC, MM, GP, MG, and SM wrote the final version of the review. The final text has been read and approved by all the authors of the review.

## FUNDING

This research was supported by grants from Sapienza Università di Roma and from Institut Pasteur (PTR-24-16).

## REFERENCES

- Ambrosi, C., Pompili, M., Scribano, D., Limongi, D., Petrucca, A., Cannavacciuolo, S., et al. (2015). The *Shigella flexneri* OspB effector: an early immunomodulator. *Int. J. Med. Microbiol.* 305, 75–84. doi: 10.1016/j.ijmm.2014.11.004
- Arbibe, L., Kim, D. W., Batsche, E., Pedron, T., Mateescu, B., Muchardt, C., et al. (2007). An injected bacterial effector targets chromatin access for transcription factor NF- $\kappa$ B to alter transcription of host genes involved in immune responses. *Nat. Immunol.* 8, 47–56. doi: 10.1038/ni1423
- Ashida, H., Ogawa, M., Kim, M., Suzuki, S., Sanada, T., Punginelli, C., et al. (2011). *Shigella* deploy multiple countermeasures against host innate immune responses. *Curr. Opin. Microbiol.* 14, 16–23. doi: 10.1016/j.mib.2010.08.014
- Ashida, H., and Sasakawa, C. (2015). *Shigella* IpaH family effectors as a versatile model for studying pathogenic bacteria. *Front. Cell. Infect. Microbiol.* 5:100. doi: 10.3389/fcimb.2015.00100
- Bando, S. Y., do Valle, G. R., Martinez, M. B., Trabulsi, L. R., and Moreira-Filho, C. A. (1998). Characterization of enteroinvasive *Escherichia coli* and *Shigella* strains by RAPD analysis. *FEMS Microbiol. Lett.* 165, 159–165. doi: 10.1111/j.1574-6968.1998.tb13141.x
- Bando, S. Y., Moreno, A. C. R., Albuquerque, J. A. T., Amhaz, J. M. K., Moreira-Filho, C. A., and Martinez, M. B. (2010). Expression of bacterial virulence factors and cytokines during in vitro macrophage infection by enteroinvasive *Escherichia coli* and *Shigella flexneri*: a comparative study. *Mem. Inst. Oswaldo Cruz* 105, 786–791. doi: 10.1590/S0074-02762010000600009
- Barbagallo, M., Di Martino, M. L., Marcocci, L., Pietrangeli, P., De Carolis, E., Casalino, M., et al. (2011). A new piece of the *Shigella* pathogenicity puzzle: spermidine accumulation by silencing of the speG gene [corrected]. *PLOS ONE* 6:e27226. doi: 10.1371/journal.pone.0027226
- Baxt, L. A., and Goldberg, M. B. (2014). Host and bacterial proteins that repress recruitment of LC3 to *Shigella* early during infection. *PLOS ONE* 9:e94653. doi: 10.1371/journal.pone.0094653
- Beutin, L., Gleier, K., Kontny, I., Echeverria, P., and Scheutz, F. (1997). Origin and characteristics of enteroinvasive strains of *Escherichia coli* (EIEC) isolated in Germany. *Epidemiol. Infect.* 118, 199–205. doi: 10.1017/S0950268897007413
- Blake, P. A., Ramos, S., MacDonald, K. L., Rassi, V., Gomes, T. A., Ivey, C., et al. (1993). Pathogen-specific risk factors and protective factors for acute diarrheal disease in urban Brazilian infants. *J. Infect. Dis.* 167, 627–632. doi: 10.1093/infdis/167.3.627
- Bliven, K. A., and Maurelli, A. T. (2012). Antivirulence genes: insights into pathogen evolution through gene loss. *Infect. Immun.* 80, 4061–4070. doi: 10.1128/IAI.00740-12
- Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., et al. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* 38, 760–771. doi: 10.1046/j.1365-2958.2000.02179.x
- Campbell-Valois, F.-X., Sachse, M., Sansonetti, P. J., and Parsot, C. (2015). Escape of actively secreting *Shigella flexneri* from ATG8/LC3-positive vacuoles formed during cell-to-cell spread is facilitated by IcsB and VirA. *mBio* 6:e02567-14. doi: 10.1128/mBio.02567-14
- Campilongo, R., Di Martino, M. L., Marcocci, L., Pietrangeli, P., Leuzzi, A., Grossi, M., et al. (2014). Molecular and functional profiling of the polyamine content in enteroinvasive *E. coli*: looking into the gap between commensal *E. coli* and harmful *Shigella*. *PLOS ONE* 9:e106589. doi: 10.1371/journal.pone.0106589
- Casalino, M., Latella, M. C., Prosseda, G., Ceccarini, P., Grimont, F., and Colonna, B. (2005). Molecular evolution of the lysine decarboxylase-defective phenotype in *Shigella sonnei*. *Int. J. Med. Microbiol.* 294, 503–512. doi: 10.1016/j.ijmm.2004.11.001
- Casalino, M., Latella, M. C., Prosseda, G., and Colonna, B. (2003). CadC is the preferential target of a convergent evolution driving enteroinvasive *Escherichia coli* toward a lysine decarboxylase-defective phenotype. *Infect. Immun.* 71, 5472–5479. doi: 10.1128/IAI.71.10.5472-5479.2003
- Casalino, M., Prosseda, G., Barbagallo, M., Iacobino, A., Ceccarini, P., Carmela Latella, M., et al. (2010). Interference of the CadC regulator in the arginine-dependent acid resistance system of *Shigella* and enteroinvasive *E. coli*. *Int. J. Med. Microbiol.* 300, 289–295. doi: 10.1016/j.ijmm.2009.10.008
- Chatterjee, B. D., and Sanyal, S. N. (1984). Is it all shigellosis? *Lancet* 2, 574. doi: 10.1016/S0140-6736(84)90783-9
- Cheasty, T., and Rowe, B. (1983). Antigenic relationships between the enteroinvasive *Escherichia coli* O antigens O28ac, O112ac, O124, O136, O143, O144, O152, and O164 and *Shigella* O antigens. *J. Clin. Microbiol.* 17, 681–684.
- Colonna, B., Casalino, M., Fradiani, P. A., Zagaglia, C., Naitza, S., Leoni, L., et al. (1995). H-NS regulation of virulence gene expression in enteroinvasive *Escherichia coli* harboring the virulence plasmid integrated into the host chromosome. *J. Bacteriol.* 177, 4703–4712. doi: 10.1128/jb.177.16.4703-4712.1995
- Croxen, M. A., and Finlay, B. B. (2009). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat. Rev. Microbiol.* 8, 26–38. doi: 10.1038/nrmicro.2265
- Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., and Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* 26, 822–880. doi: 10.1128/CMR.00022-13



- Dagberg, B., and Uhlin, B. E. (1992). Regulation of virulence-associated plasmid genes in enteroinvasive *Escherichia coli*. *J. Bacteriol.* 174, 7606–7612. doi: 10.1128/jb.174.23.7606-7612.1992
- Di Martino, M. L., Campilongo, R., Casalino, M., Micheli, G., Colonna, B., and Prosseda, G. (2013a). Polymines: emerging players in bacteria-host interactions. *Int. J. Med. Microbiol.* 303, 484–491. doi: 10.1016/j.ijmm.2013.06.008
- Di Martino, M. L., Falconi, M., Micheli, G., Colonna, B., and Prosseda, G. (2016a). The multifaceted activity of the VirF regulatory protein in the *Shigella* lifestyle. *Front. Mol. Biosci.* 3:61. doi: 10.3389/fmolb.2016.00061
- Di Martino, M. L., Fioravanti, R., Barbabella, G., Prosseda, G., Colonna, B., and Casalino, M. (2013b). Molecular evolution of the nicotinic acid requirement within the *Shigella*/EIEC pathotype. *Int. J. Med. Microbiol.* 303, 651–661. doi: 10.1016/j.ijmm.2013.09.007
- Di Martino, M. L., Romilly, C., Wagner, E. G. H., Colonna, B., and Prosseda, G. (2016b). One gene and two proteins: a leaderless mRNA supports the translation of a shorter form of the *Shigella* VirF regulator. *mBio* 7:e01860-16. doi: 10.1128/mBio.01860-16
- Dobrindt, U., Chowdary, M. G., Krumbholz, G., and Hacker, J. (2010). Genome dynamics and its impact on evolution of *Escherichia coli*. *Med. Microbiol. Immunol.* 199, 145–154. doi: 10.1007/s00430-010-0161-2
- Dong, N., Zhu, Y., Lu, Q., Hu, L., Zheng, Y., and Shao, F. (2012). Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell* 150, 1029–1041. doi: 10.1016/j.cell.2012.06.050
- Dorman, C. J. (2004). H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.* 2, 391–400. doi: 10.1038/nrmicro883
- DuPont, H. L., Formal, S. B., Hornick, R. B., Snyder, M. J., Libonati, J. P., Sheahan, D. G., et al. (1971). Pathogenesis of *Escherichia coli* diarrhea. *N. Engl. J. Med.* 285, 1–9. doi: 10.1056/NEJM197107012850101
- Echeverria, P., Sethabutr, O., Serichantalergs, O., Lexomboon, U., and Tamura, K. (1992). *Shigella* and enteroinvasive *Escherichia coli* infections in households of children with dysentery in Bangkok. *J. Infect. Dis.* 165, 144–147. doi: 10.1093/infdis/165.1.144
- Edwards, P. R., and Ewing, W. H. (1986). *Edwards and Ewing's Identification of Enterobacteriaceae*, 4th Edn. New York, NY: Elsevier Science.
- Escher, M., Scavia, G., Morabito, S., Tozzoli, R., Maugliani, A., Cantoni, S., et al. (2014). A severe foodborne outbreak of diarrhoea linked to a canteen in Italy caused by enteroinvasive *Escherichia coli*, an uncommon agent. *Epidemiol. Infect.* 142, 2559–2566. doi: 10.1017/S0950268814000181
- Escobar-Páramo, P., Giudicelli, C., Parsot, C., and Denamur, E. (2003). The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J. Mol. Evol.* 57, 140–148. doi: 10.1007/s00239-003-2460-3
- Ewing, W. H., and Gravatt, J. L. (1947). *Shigella* types encountered in the mediterranean area. *J. Bacteriol.* 53, 191–195.
- Farmer, J. J., Davis, B. R., Hickman-Brenner, F. W., McWhorter, A., Huntley-Carter, G. P., Asbury, M. A., et al. (1985). Biochemical identification of new species and biogroups of Enterobacteriaceae isolated from clinical specimens. *J. Clin. Microbiol.* 21, 46–76.
- Faundez, G., Figueroa, G., Troncoso, M., and Cabello, F. C. (1988). Characterization of enteroinvasive *Escherichia coli* strains isolated from children with diarrhea in Chile. *J. Clin. Microbiol.* 26, 928–932.
- Fernandez, I. M., Silva, M., Schuch, R., Walker, W. A., Siber, A. M., Maurelli, A. T., et al. (2001). Cadaverine prevents the escape of *Shigella flexneri* from the phagolysosome: a connection between bacterial dissemination and neutrophil transepithelial signaling. *J. Infect. Dis.* 184, 743–753. doi: 10.1086/323035
- Formal, S. B., and Hornick, R. B. (1978). Invasive *Escherichia coli*. *J. Infect. Dis.* 137, 641–644. doi: 10.1093/infdis/137.5.641
- Giangrossi, M., Prosseda, G., Tran, C. N., Brandi, A., Colonna, B., and Falconi, M. (2010). A novel antisense RNA regulates at transcriptional level the virulence gene *icsA* of *Shigella flexneri*. *Nucleic Acids Res.* 38, 3362–3375. doi: 10.1093/nar/gkq025
- Gomes, T. A., Toledo, M. R., Trabulsi, L. R., Wood, P. K., and Morris, J. G. (1987). DNA probes for identification of enteroinvasive *Escherichia coli*. *J. Clin. Microbiol.* 25, 2025–2027.
- Gomes, T. A. T., Elias, W. P., Scaletsky, I. C. A., Guth, B. E. C., Rodrigues, J. F., Piazza, R. M. F., et al. (2016). Diarrheagenic *Escherichia coli*. *Braz. J. Microbiol.* 47, 3–30. doi: 10.1016/j.bjm.2016.10.015
- Hale, T. L., Oaks, E. V., and Formal, S. B. (1985). Identification and antigenic characterization of virulence-associated, plasmid-coded proteins of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Infect. Immun.* 50, 620–629.
- Hale, T. L., Sansonetti, P. J., Schad, P. A., Austin, S., and Formal, S. B. (1983). Characterization of virulence plasmids and plasmid-associated outer membrane proteins in *Shigella flexneri*, *Shigella sonnei*, and *Escherichia coli*. *Infect. Immun.* 40, 340–350.
- Harris, J. R., Wachsmuth, I. K., Davis, B. R., and Cohen, M. L. (1982). High-molecular-weight plasmid correlates with *Escherichia coli* enteroinvasiveness. *Infect. Immun.* 37, 1295–1298.
- Hazen, T. H., Leonard, S. R., Lampel, K. A., Lacher, D. W., Maurelli, A. T., and Rasko, D. A. (2016). Investigating the relatedness of enteroinvasive *Escherichia coli* to other *E. coli* and *Shigella* isolates by using comparative genomics. *Infect. Immun.* 84, 2362–2371. doi: 10.1128/IAI.00350-16
- Johnson, T. J., and Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 73, 750–774. doi: 10.1128/MMBR.00015-09
- Kaper, J. B., Nataro, J. P., and Mobley, H. L. T. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818
- Killackey, S. A., Sorbara, M. T., and Girardin, S. E. (2016). Cellular aspects of *Shigella* pathogenesis: focus on the manipulation of host cell processes. *Front. Cell. Infect. Microbiol.* 6:38. doi: 10.3389/fcimb.2016.00038
- Kim, D. W., Lenzen, G., Page, A.-L., Legrain, P., Sansonetti, P. J., and Parsot, C. (2005). The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14046–14051. doi: 10.1073/pnas.0504466102
- Kopecko, D. J., Baron, L. S., and Buysse, J. (1985). Genetic determinants of virulence in *Shigella* and dysenteric strains of *Escherichia coli*: their involvement in the pathogenesis of dysentery. *Curr. Top. Microbiol. Immunol.* 118, 71–95. doi: 10.1007/978-3-642-70586-1\_5
- Lan, R., Alles, M. C., Donohoe, K., Martinez, M. B., and Reeves, P. R. (2004). Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infect. Immun.* 72, 5080–5088. doi: 10.1128/IAI.72.9.5080-5088.2004
- Lan, R., Lumb, B., Ryan, D., and Reeves, P. R. (2001). Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. *Infect. Immun.* 69, 6303–6309. doi: 10.1128/IAI.69.10.6303-6309.2001
- Lan, R., and Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect.* 4, 1125–1132. doi: 10.1016/S1286-4579(02)01637-4
- Leuzzi, A., Di Martino, M. L., Campilongo, R., Falconi, M., Barbagallo, M., Marocchi, L., et al. (2015). Multifactor regulation of the MdtJ polyamine transporter in *Shigella*. *PLOS ONE* 10:e0136744. doi: 10.1371/journal.pone.0136744
- Leuzzi, A., Grossi, M., Di Martino, M. L., Pasqua, M., Micheli, G., Colonna, B., et al. (2017). Role of the SRRz/Rz1 lambdoid lysis cassette in the pathoadaptive evolution of *Shigella*. *Int. J. Med. Microbiol.* 307, 268–275. doi: 10.1016/j.ijmm.2017.03.002
- Levine, M. M., Ferreccio, C., Prado, V., Cayazzo, M., Abrego, P., Martinez, J., et al. (1993). Epidemiologic studies of *Escherichia coli* diarrheal infections in a low socioeconomic level peri-urban community in Santiago. *Chile. Am. J. Epidemiol.* 138, 849–869. doi: 10.1093/oxfordjournals.aje.a116788
- Makino, S., Sasakawa, C., and Yoshikawa, M. (1988). Genetic relatedness of the basic replicon of the virulence plasmid in shigellae and enteroinvasive *Escherichia coli*. *Microb. Pathog.* 5, 267–274. doi: 10.1016/0882-4010(88)90099-X
- Manolov, D. G. (1959). A new type of the genus *Shigella* – “*Shigella* 13”. *J. Hyg. Epidemiol. Microbiol. Immunol.* 3, 184–190.
- Marier, R., Wells, J., Swanson, R., Callahan, W., and Mehlman, I. (1973). An outbreak of enteropathogenic *Escherichia coli* foodborne disease traced to imported French cheese. *Lancet* 302, 1376–1378. doi: 10.1016/S0140-6736(73)93355-7
- Matsushita, S., Yamada, S., Kai, A., and Kudoh, Y. (1993). Invasive strains of *Escherichia coli* belonging to serotype O121:NM. *J. Clin. Microbiol.* 31, 3034–3035.
- Mattock, E., and Blocker, A. J. (2017). How do the virulence factors of *Shigella* work together to cause disease? *Front. Cell. Infect. Microbiol.* 7:64. doi: 10.3389/fcimb.2017.00064

- Michelacci, V., Prosseda, G., Maugliani, A., Tozzoli, R., Sanchez, S., Herrera-León, S., et al. (2016). Characterization of an emergent clone of enteroinvasive *Escherichia coli* circulating in Europe. *Clin. Microbiol. Infect.* 22, 287.e11–287.e19. doi: 10.1016/j.cmi.2015.10.025
- Moreno, A. C. R., Ferreira, L. G., and Martinez, M. B. (2009). Enteroinvasive *Escherichia coli* vs. *Shigella flexneri*: how different patterns of gene expression affect virulence. *FEMS Microbiol. Lett.* 301, 156–163. doi: 10.1111/j.1574-6968.2009.01815.x
- Nakata, N., Tobe, T., Fukuda, I., Suzuki, T., Komatsu, K., Yoshikawa, M., et al. (1993). The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the ompT and kcpA loci. *Mol. Microbiol.* 9, 459–468. doi: 10.1111/j.1365-2958.1993.tb01707.x
- Newitt, S., MacGregor, V., Robbins, V., Bayliss, L., Chattaway, M. A., Dallman, T., et al. (2016). Two linked enteroinvasive *Escherichia coli* outbreaks, Nottingham, UK, June 2014. *Emerg. Infect. Dis.* 22, 1178–1184. doi: 10.3201/eid2207.152080
- O'Brien, A. D., Gentry, M. K., Thompson, M. R., Doctor, B. P., Gemski, P., and Formal, S. B. (1979). Shigellosis and *Escherichia coli* diarrhea: relative importance of invasive and toxigenic mechanisms. *Am. J. Clin. Nutr.* 32, 229–233.
- Ogawa, M., Yoshimori, T., Suzuki, T., Sagara, H., Mizushima, N., and Sasakawa, C. (2005). Escape of intracellular *Shigella* from autophagy. *Science* 307, 727–731. doi: 10.1126/science.1106036
- Orskov, I., Wachsmuth, I. K., Taylor, D. N., Echeverria, P., Rowe, B., Sakazaki, R., et al. (1991). Two new *Escherichia coli* O groups: O172 from “Shiga-like” toxin II-producing strains (EHEC) and O173 from enteroinvasive *E. coli* (EIEC). *APMIS* 99, 30–32. doi: 10.1111/j.1699-0463.1991.tb05114.x
- Parsot, C. (2005). *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. *FEMS Microbiol. Lett.* 252, 11–18. doi: 10.1016/j.femsle.2005.08.046
- Pavlovic, M., Luze, A., Konrad, R., Berger, A., Sing, A., Busch, U., et al. (2011). Development of a duplex real-time PCR for differentiation between *E. coli* and *Shigella* spp. *J. Appl. Microbiol.* 110, 1245–1251. doi: 10.1111/j.1365-2672.2011.04973.x
- Pettengill, E. A., Pettengill, J. B., and Binet, R. (2015). Phylogenetic analyses of *Shigella* and enteroinvasive *Escherichia coli* for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. *Front. Microbiol.* 6:1573. doi: 10.3389/fmicb.2015.01573
- Pettengill, E. A., Pettengill, J. B., and Binet, R. (2016). Phylogenetic analyses of *Shigella* and enteroinvasive *Escherichia coli* for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. *Front. Microbiol.* 6:1573. doi: 10.3389/fmicb.2015.01573
- Prosseda, G., Carmela Latella, M., Barbagallo, M., Nicoletti, M., Al Kassas, R., Casalino, M., et al. (2007). The two-faced role of cad genes in the virulence of pathogenic *Escherichia coli*. *Res. Microbiol.* 158, 487–493. doi: 10.1016/j.resmic.2007.05.001
- Prosseda, G., Di Martino, M. L., Campilongo, R., Fioravanti, R., Micheli, G., Casalino, M., et al. (2012). Shedding of genes that interfere with the pathogenic lifestyle: the *Shigella* model. *Res. Microbiol.* 163, 399–406. doi: 10.1016/j.resmic.2012.07.004
- Prosseda, G., Falconi, M., Giangrossi, M., Gualerzi, C. O., Micheli, G., and Colonna, B. (2004). The virF promoter in *Shigella*: more than just a curved DNA stretch. *Mol. Microbiol.* 51, 523–537. doi: 10.1046/j.1365-2958.2003.03848.x
- Prosseda, G., Falconi, M., Nicoletti, M., Casalino, M., Micheli, G., and Colonna, B. (2002). Histone-like proteins and the *Shigella* invasivity regulon. *Res. Microbiol.* 153, 461–468. doi: 10.1016/S0923-2508(02)01346-3
- Prosseda, G., Latella, M. C., Casalino, M., Nicoletti, M., Michienzi, S., and Colonna, B. (2006). Plasticity of the P<sub>junc</sub> promoter of ISEc11, a new insertion sequence of the IS1111 family. *J. Bacteriol.* 188, 4681–4689. doi: 10.1128/JB.00332-06
- Prunier, A.-L., Schuch, R., Fernandez, R. E., Mumy, K. L., Kohler, H., McCormick, B. A., et al. (2007). nadA and nadB of *Shigella flexneri* 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of *Shigella* pathogenicity. *Microbiology* 153, 2363–2372. doi: 10.1099/mic.0.2007/006916-0
- Pupo, G. M., Lan, R., and Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10567–10572. doi: 10.1073/pnas.180094797
- Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., et al. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190, 6881–6893. doi: 10.1128/JB.00619-08
- Reiterer, V., Grossniklaus, L., Tschon, T., Kasper, C. A., Sorg, I., and Arrieumerlou, C. (2011). *Shigella flexneri* type III secreted effector OspF reveals new crosstalks of proinflammatory signaling pathways during bacterial infection. *Cell. Signal.* 23, 1188–1196. doi: 10.1016/j.cellsig.2011.03.006
- Rowe, B., Gross, R. J., and Woodroof, D. P. (1977). Proposal to recognise serovar 145/46 (synonyms: 147, *Shigella* 13, *Shigella sofia*, and *Shigella manolovii*) as a New *Escherichia coli* O group, O164. *Int. J. Syst. Bacteriol.* 27, 15–18. doi: 10.1099/00207713-27-1-15
- Sahl, J. W., Morris, C. R., Emberger, J., Fraser, C. M., Ochieng, J. B., Juma, J., et al. (2015). Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J. Clin. Microbiol.* 53, 951–960. doi: 10.1128/JCM.03527-14
- Sakazaki, R., Tamura, K., and Saito, M. (1967). Enteropathogenic *Escherichia coli* associated with diarrhea in children and adults. *Jpn. J. Med. Sci. Biol.* 20, 387–399. doi: 10.7883/yoken1952.20.387
- Sanchez-Villamil, J., Tapia-Pastrana, G., and Navarro-Garcia, F. (2016). Pathogenic lifestyles of *E. coli* pathotypes in a standardized epithelial cell model influence inflammatory signaling pathways and cytokines secretion. *Front. Cell. Infect. Microbiol.* 6:120. doi: 10.3389/fcimb.2016.00120
- Sansonetti, P. J., Hale, T. L., Dammin, G. J., Kapfer, C., Collins, H. H., and Formal, S. B. (1983). Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. *Infect. Immun.* 39, 1392–1402.
- Sansonetti, P. J., Kopecko, D. J., and Formal, S. B. (1982). Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect. Immun.* 35, 852–860.
- Schroeder, G. N., and Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin. Microbiol. Rev.* 21, 134–156. doi: 10.1128/CMR.00032-07
- Scribano, D., Petrucca, A., Pompili, M., Ambrosi, C., Bruni, E., Zagaglia, C., et al. (2014). Polar localization of PhoN2, a periplasmic virulence-associated factor of *Shigella flexneri*, is required for proper IcsA exposition at the old bacterial pole. *PLOS ONE* 9:e90230. doi: 10.1371/journal.pone.0090230
- Silva, R. M., Toledo, M. R., and Trabulsi, L. R. (1980). Biochemical and cultural characteristics of invasive *Escherichia coli*. *J. Clin. Microbiol.* 11, 441–444.
- Small, P. L., and Falkow, S. (1988). Identification of regions on a 230-kilobase plasmid from enteroinvasive *Escherichia coli* that are required for entry into HEp-2 cells. *Infect. Immun.* 56, 225–229.
- Svenungsson, B., Lagergren, A., Ekwall, E., Evengård, B., Hedlund, K. O., Kärrnall, A., et al. (2000). Enteropathogens in adult patients with diarrhea and healthy control subjects: a 1-year prospective study in a Swedish clinic for infectious diseases. *Clin. Infect. Dis.* 30, 770–778. doi: 10.1086/313770
- Takeuchi, O., and Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell* 140, 805–820. doi: 10.1016/j.cell.2010.01.022
- Taylor, D. N., Echeverria, P., Sethabutr, O., Pitarangsi, C., Leksomboon, U., Blacklow, N. R., et al. (1988). Clinical and microbiologic features of *Shigella* and enteroinvasive *Escherichia coli* infections detected by DNA hybridization. *J. Clin. Microbiol.* 26, 1362–1366.
- Toledo, M. R., and Trabulsi, L. R. (1983). Correlation between biochemical and serological characteristics of *Escherichia coli* and results of the Serény test. *J. Clin. Microbiol.* 17, 419–421.
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLOS Genet.* 5:e1000344. doi: 10.1371/journal.pgen.1000344
- Tozzoli, R., and Scheutz, F. (2014). “Diarrheagenic *Escherichia coli* infections in humans,” in *Pathogenic Escherichia coli: Molecular and Cellular Microbiology*, ed. S. Morabito (Norfolk: Caister Academic Press).
- Trabulsi, L. R., Fernandes, M. R., and Zuliani, M. E. (1967). New intestinal bacteria pathogenic to man. *Rev. Inst. Med. Trop. Sao Paulo* 9, 31–39.
- Tran, C. N., Giangrossi, M., Prosseda, G., Brandi, A., Di Martino, M. L., Colonna, B., et al. (2011). A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene

- icsA of *Shigella flexneri*. *Nucleic Acids Res.* 39, 8122–8134. doi: 10.1093/nar/gkr521
- Turner, S. A., Luck, S. N., Sakellaris, H., Rajakumar, K., and Adler, B. (2003). Molecular epidemiology of the SRL pathogenicity island. *Antimicrob. Agents Chemother.* 47, 727–734. doi: 10.1128/AAC.47.2.727-734.2003
- van den Beld, M. J. C., and Reubsat, F. A. G. (2012). Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 899–904. doi: 10.1007/s10096-011-1395-7
- van Elsas, J. D., Semenov, A. V., Costa, R., and Trevors, J. T. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5, 173–183. doi: 10.1038/ismej.2010.80
- Venkatesan, M. M., Goldberg, M. B., Rose, D. J., Grotbeck, E. J., Burland, V., and Blattner, F. R. (2001). Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infect. Immun.* 69, 3271–3285. doi: 10.1128/IAI.69.5.3271-3285.2001
- Vieira, N., Bates, S. J., Solberg, O. D., Ponce, K., Howsmon, R., Cevallos, W., et al. (2007). High prevalence of enteroinvasive *Escherichia coli* isolated in a remote region of northern coastal Ecuador. *Am. J. Trop. Med. Hyg.* 76, 528–533.
- Voerres, S., Redey, B., and Csizmazia, F. (1964). Antigenic structure of a new enteropathogenic *E. coli* strain. *Acta Microbiol. Acad. Sci. Hung.* 11, 125–129.
- Wanger, A. R., Murray, B. E., Echeverria, P., Mathewson, J. J., and DuPont, H. L. (1988). Enteroinvasive *Escherichia coli* in travelers with diarrhea. *J. Infect. Dis.* 158, 640–642. doi: 10.1093/infdis/158.3.640
- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., et al. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60, 1136–1151. doi: 10.1111/j.1365-2958.2006.05172.x
- Zagaglia, C., Casalino, M., Colonna, B., Conti, C., Calconi, A., and Nicoletti, M. (1991). Virulence plasmids of enteroinvasive *Escherichia coli* and *Shigella flexneri* integrate into a specific site on the host chromosome: integration greatly reduces expression of plasmid-carried virulence genes. *Infect. Immun.* 59, 792–799.
- Zuo, G., Xu, Z., and Hao, B. (2013). *Shigella* strains are not clones of *Escherichia coli* but sister species in the genus *Escherichia*. *Genomics Proteomics Bioinformatics* 11, 61–65. doi: 10.1016/j.gpb.2012.11.002
- Zurawski, D. V., Mumy, K. L., Faherty, C. S., McCormick, B. A., and Maurelli, A. T. (2009). *Shigella flexneri* type III secretion system effectors OspB and OspF target the nucleus to downregulate the host inflammatory response via interactions with retinoblastoma protein. *Mol. Microbiol.* 71, 350–368. doi: 10.1111/j.1365-2958.2008.06524.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pasqua, Michelacci, Di Martino, Tozzoli, Grossi, Colonna, Morabito and Prosseda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Environmental Origin of the Genus *Bordetella*

Illiasou Hamidou Soumana<sup>1,2†</sup>, Bodo Linz<sup>2,3\*†</sup> and Eric T. Harvill<sup>1,2,3\*</sup>

<sup>1</sup> Department of Infectious Diseases, University of Georgia, Athens, GA, USA, <sup>2</sup> Center for Vaccines and Immunology, University of Georgia, Athens, GA, USA, <sup>3</sup> Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
The University of Texas Medical  
Branch at Galveston, USA

### Reviewed by:

Nikolai Ravin,  
Research Center for Biotechnology  
(RAS), Russia  
Louise Temple,  
James Madison University, USA

### \*Correspondence:

Bodo Linz  
bodo.linz@uga.edu  
Eric T. Harvill  
harvill@uga.edu

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 20 November 2016

Accepted: 05 January 2017

Published: 24 January 2017

### Citation:

Hamidou Soumana I, Linz B and  
Harvill ET (2017) Environmental Origin  
of the Genus *Bordetella*.  
Front. Microbiol. 8:28.  
doi: 10.3389/fmicb.2017.00028

Members of the genus *Bordetella* include human and animal pathogens that cause a variety of respiratory infections, including whooping cough in humans. Despite the long known ability to switch between a within-animal and an extra-host lifestyle under laboratory growth conditions, no extra-host niches of pathogenic *Bordetella* species have been defined. To better understand the distribution of *Bordetella* species in the environment, we probed the NCBI nucleotide database with the 16S ribosomal RNA (16S rRNA) gene sequences from pathogenic *Bordetella* species. Bacteria of the genus *Bordetella* were frequently found in soil, water, sediment, and plants. Phylogenetic analyses of their 16S rRNA gene sequences showed that *Bordetella* recovered from environmental samples are evolutionarily ancestral to animal-associated species. Sequences from environmental samples had a significantly higher genetic diversity, were located closer to the root of the phylogenetic tree and were present in all 10 identified sequence clades, while only four sequence clades possessed animal-associated species. The pathogenic bordetellae appear to have evolved from ancestors in soil and/or water. We show that, despite being animal-adapted pathogens, *Bordetella bronchiseptica*, and *Bordetella hinzii* have preserved the ability to grow and proliferate in soil. Our data implicate soil as a probable environmental origin of *Bordetella* species, including the animal-pathogenic lineages. Soil may further constitute an environmental niche, allowing for persistence and dissemination of the bacterial pathogens. Spread of pathogenic bordetellae from an environmental reservoir such as soil may potentially explain their wide distribution as well as frequent disease outbreaks that start without an obvious infectious source.

**Keywords:** *Bordetella*, environmental strains, ecological niches, extra-host adaptation, environmental origin

## INTRODUCTION

Bacteria of the genus *Bordetella* are of primary importance in human and veterinary medicine because of their ability to colonize the respiratory tract, causing a wide range of pulmonary and bronchial infections. The common human- and animal-adapted pathogens *B. pertussis*, *B. parapertussis*, and *B. bronchiseptica* are known as the “classical” *Bordetella* species. *B. pertussis* is a strictly human pathogen, but *B. parapertussis* consists of two lineages, one infecting humans and the other infecting sheep (Mattoo and Cherry, 2005). In contrast to these examples of adaptation to a single host, *B. bronchiseptica* colonizes a variety of animals and even humans (Register et al., 2015), resulting in a broad array of respiratory diseases, from asymptomatic colonization to lethal



pneumonia (Goodnow, 1980). Phylogenetic analyses (Musser et al., 1986; Diavatopoulos et al., 2005) and genome comparisons (Parkhill et al., 2003) have revealed that *B. pertussis* and *B. parapertussis* represent human-adapted forms of *B. bronchiseptica* that have evolved independently from a *B. bronchiseptica*-like ancestor. The genus also contains a number of additional, more recently described species. For example, *B. avium* (Kerstens et al., 1984) causes respiratory disease in birds. *B. hinzii* (Vandamme et al., 1995) is generally regarded as a non-pathogenic colonizer of the respiratory tract of poultry but some strains were shown to cause disease in turkeys when experimentally inoculated (Register and Kunkle, 2009). Meanwhile, the closely related species *B. pseudohinzii* colonizes laboratory mice (Ivanov et al., 2015, 2016). *B. holmesii* (Weyant et al., 1995) causes pertussis-like disease and septicemia in humans (Shepard et al., 2004), and *B. bronchialis*, *B. flabilis*, and *B. sputigena* (Vandamme et al., 2015) were also isolated from human respiratory specimens. In contrast to other bordetellae, *B. trematum* (Vandamme et al., 1996) and *B. ansorpii* (Ko et al., 2005) are not associated with respiratory problems but were isolated from human wound infection.

*B. petrii*, a species originally isolated from a dechlorinating bioreactor enriched by river sediment, represents the first described environmental species within the *Bordetella* genus (von Wintzingerode et al., 2001). *B. petrii* strains were also found in marine sponges (Wang et al., 2007), grass root consortia (Wang et al., 2007), and in other environmental samples as members of microbial communities involved in the degradation of aromatic hydrocarbons, such as benzenes (Bianchi et al., 2005; Wang et al., 2007). In apparent conflict with its environmental source, the *B. petrii* genome contains genes that allow for the synthesis and secretion of factors specifically associated with the virulence of the pathogenic *Bordetella* sp., including the BvgAS master regulon and filamentous hemagglutinin (Gross et al., 2008). In addition to these environmental sources, *B. petrii* was also isolated from immunocompromised patients with ear infection, cystic fibrosis and chronic pulmonary disease (Fry et al., 2005; Biederman et al., 2015; Nagata et al., 2015), suggesting broad adaptability of this bacterial species to both environmental conditions and as an opportunistic pathogen of humans and possibly other animals.

Other *Bordetella* species have been obtained from environments not associated with animal hosts. Ten different bacterial strains were cultured from cotton swabs taken from the plaster wall surface of 1300-year-old mural paintings inside the stone chamber of the Takamatsuzuka Tomb, an ancient circular burial mound in Japan. Taxonomic classification of these isolates revealed three novel species that were then named *B. muralis*, *B. tumulicola*, and *B. tumbae* (Tazato et al., 2015). Isolation of the bacteria from the paintings, but not from the surrounding stone walls, suggested that these species might be involved in the observed biodeterioration of the colorful paintings (Kigawa et al., 2013).

According to their 16S rRNA gene sequences, other environmental bacteria from soil also belong to the genus *Bordetella*. Interestingly, the majority of those samples originated from contaminated sites such as soil polluted with chlorinated

benzenes (Wang et al., 2007), from arctic soils contaminated with polycyclic aromatic hydrocarbons such as oil, diesel fuel or tar (Eriksson et al., 2003), from the sediment of a municipal wastewater plant (Nisola et al., 2010) and from arsenic polluted soils (Cavalca et al., 2010; Bachate et al., 2012). All these observations suggest that members of the *Bordetella* genus may have the potential for biodegradation of a great variety of organic compounds.

Although these anecdotal findings suggest that members of the *Bordetella* genus may be found in nature, there is currently no systematic analysis of the occurrence of *Bordetella* outside human or animal hosts, and the potential impact of environmental isolates on human and animal health is uncertain. Environmental niches of pathogenic *Bordetella* species have been proposed but not identified. Yet, the ability of *Bordetella* to survive and persist outside mammalian hosts would allow for its greater dissemination and persistence, and could contribute to a wide distribution of infections and disease, often without an obvious infectious source.

Here, we search the NCBI nucleotide database for 16S ribosomal RNA gene sequences of *Bordetella*-like microorganisms from various environments and compare them to those of the described species, including the classical bordetellae, to determine their phylogenetic relatedness. We identified 10 clades of related strains, all of which contained samples isolated from environmental sources, though only four clades also contained sequences from animal-associated species. Sequences from environmental samples had a significantly higher genetic diversity and were located closer to the root of the phylogenetic tree than those from animal-associated isolates, suggesting an environmental origin of the genus *Bordetella*. In addition, we show that the animal-adapted pathogens *B. bronchiseptica* and *B. hinzii* grow efficiently in soil extract, indicating that diverse pathogenic bordetellae may have retained the ability to proliferate in the environment.

## MATERIALS AND METHODS

### Search for *Bordetella* 16S rRNA Gene Sequences in the NCBI nt Database

The 16S ribosomal gene sequences of *Bordetella hinzii* strain LMG 13501 (GenBank accession number NR\_027537.1); *Bordetella holmesii* strain ATCC 51541 (NR\_121717.1); and *Bordetella pertussis* strain Tohama I (AF142326.1) were each used as queries for BLAST search (blastn) against the NCBI nr/nt database using the default search parameters with a hitlist size of 5000. From the numerous hits, we excluded sequences of isolates from the known species that are mentioned in the introduction and selected only those that showed higher percentage of similarity to known *Bordetella* species than to bacteria from any other genus, including *Achromobacter*. As a control, we ran blastn searches with each of the identified sequences against the NCBI nr/nt database to remove potential false positives. The remaining sequences, all of which were from bacteria obtained from environmental sources, were considered for further analysis. All three searches using 16S rRNA sequences of

*B. pertussis*, *B. hinzii*, and *B. holmesii* as queries, respectively, gave consistent results. The accession numbers were then explored for details on sample source, year of isolation, and associated publications. Most sequences were described as *Bordetella* sp. in the gene description, but some were designated as “uncultured bacterium.”

## Phylogenetic Analysis and Tree Construction

All 16S rRNA sequences were aligned in Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), and the alignment was checked manually for consistency. Only sequences containing a 1376 bp gene fragment (near full-length) were used for further analyses. In order to identify the closely related species of environmental *Bordetella* isolates, the 16S ribosomal RNA gene sequences of members of 16 named *Bordetella* species were used as references; namely *B. pertussis* Tohama I, *B. parapertussis* BPP5, *B. bronchiseptica* RB50, *B. avium* 197N, *B. hinzii* LMG 13501, *B. pseudohinzii* 8-296-03, *B. holmesii* ATCC 51541, *B. trematum* DSM 11334, *B. ansorpii* SMC-8986, *B. bronchialis* LMG 28640, *B. sputigena* LMG 28641, *B. flabilis* LMG 28642, *B. petrii* Se-1111R, *B. muralis* T6220-3-2b, *B. tumulicola* T6517-1-4b, and *B. tumbae* T6713-1-3b (Table 1). The 16S rRNA gene sequences of *Burkholderia pseudomallei* NCTC13179 and *Ralstonia solanacearum* YP-01 were used as outgroups. The aligned and trimmed sequences (one per unique sequence) were used to generate a Neighbor-joining tree using the Maximum Composite Likelihood algorithm in MEGA (Tamura et al., 2007), and bootstrap support was estimated running 100,000 replications. Nucleotide diversity ( $\Pi$ ) within environmental samples and within animal-associated samples were estimated in DnaSP (Librado and Rozas, 2009),

and 95% confidence limits ( $\Pi_{95}$ ) were estimated using an online confidence limit calculator (<https://www.allto.co.uk/tools/statistic-calculators/confidence-interval-for-mean-calculator/>).

## Soil Sample Collection and *Bordetella* Growth in Soil Extract

Soils were sampled in April 2016 at two random sites in State College, Pennsylvania, near a suburban park (40°48'40.7" N 77°53'06.1" W and 40°48'38.2" N 77°53'04.2" W). Each sample was collected to a depth of 20 cm and thoroughly mixed. Fifty grams of each soil sample (100 g total) was placed in a bottle which was filled to 500 ml with sterile PBS. The sample was homogenized by shaking for 10 min, then left to settle for 1 h at room temperature and carefully decanted. The soil-PBS suspension was filter sterilized. Single colonies of *B. bronchiseptica* strain RB50, *B. hinzii* strain L60, and *B. petrii* strain DSMZ12804 were picked from Bordet-Gengou (BG) agar (Difco) plates supplemented with 10% defibrinated Sheep's blood (HemoStat Laboratories, Dixon, CA, USA) and were cultured in liquid Stainer-Scholte medium (Stainer and Scholte, 1970) overnight at 37°C. The *Bordetella* inocula were prepared as follows. The cultures were centrifuged, resuspended in 1 ml PBS, and the optical density (OD<sub>600</sub>) was determined. Following five consecutive 10-fold dilutions in 1 ml PBS, 100  $\mu$ l (=10<sup>6</sup>-fold dilution) containing ~150 (*B. petrii*) or 240 bacterial cells (*B. hinzii* or *B. bronchiseptica*) were added to 5 ml of the soil extract resulting in starting concentrations of ~30 bacterial cells/ml (*B. petrii*) and 48 bacterial cells/ml (*B. hinzii*, *B. bronchiseptica*). Bacterial numbers were determined by plating an aliquot of each inoculum. The culture tubes were incubated at room temperature (25°C) with shaking. After 24, 48, and 72 h, 100  $\mu$ l of each culture was plated on BG agar supplemented with

**TABLE 1 | Reference strains of named *Bordetella* species.**

Strain name	References	GenBank accession number	16S rRNA sequence length (bp)
<i>Burkholderia pseudomallei</i> NCTC 13179	Johnson et al., 2015	CP003976.1	1487
<i>Ralstonia solanacearum</i> YP-01	NCBI	FJ494776.1	1500
<i>Bordetella avium</i> 197N	Sebahia et al., 2006	NR_074639.1	1487
<i>Bordetella bronchiseptica</i> RB50	Parkhill et al., 2003	BX640447.1	1487
<i>Bordetella hinzii</i> LMG 13501	Kattar et al., 2000	NR_027537.1	1487
<i>Bordetella parapertussis</i> BPP5	Park et al., 2012	HE965803.1	1489
<i>Bordetella holmesii</i> ATCC 51541	NCBI	NR_121717.1	1487
<i>Bordetella pertussis</i> Tohama I	Parkhill et al., 2003	AF142326.1	1487
<i>Bordetella trematum</i> DSM 11334	von Wintzingerode et al., 2001	NR_025404.1	1521
<i>Bordetella flabilis</i> LMG 28642	Vandamme et al., 2015	EU082162.1	1376
<i>Bordetella bronchialis</i> LMG 28640	Vandamme et al., 2015	EU082135.1	1416
<i>Bordetella sputigena</i> LMG 28641	Vandamme et al., 2015	KF601914.1	1376
<i>Bordetella ansorpii</i> SMC-8986	Ko et al., 2005	AY594190.1	1424
<i>Bordetella pseudohinzii</i> 8-296-03	Ivanov et al., 2016	JHEP02000033.1	1542
<i>Bordetella petrii</i> DSMZ12804	Gross et al., 2008	NC_010170	1487
<i>Bordetella muralis</i> T6220-3-2b	Tazato et al., 2015	LC053647.1	1456
<i>Bordetella tumbae</i> T6713-1-3b	Tazato et al., 2015	LC053656.1	1456
<i>Bordetella tumulicola</i> T6517-1-4b	Tazato et al., 2015	LC053650.1	1456

10% defibrinated sheep's blood to determine bacterial numbers. Each experiment was carried out in triplicate. The mean and  $\pm$  standard error as well as analysis of variance (ANOVA) were conducted using Graphpad Prism version 6.04. The bacterial doubling time was calculated by the formula: doubling time =  $\ln(2)/\ln(N(t)/N(0))/t$ , where  $N(t)$  is the number of bacterial cells at time  $t$ ,  $N(0)$  is the number of bacteria at time 0 and  $t$  is the time in hours.

## RESULTS

### *Bordetella* in the Environment

We mined the NCBI nucleotide databases for *Bordetella* spp. 16S rRNA gene sequences. The search resulted in a total of 71 *Bordetella* spp. 16S rRNA gene sequences (Table 2) in addition to those from the named species (Table 1) *B. bronchiseptica*, *B. parapertussis*, *B. pertussis*, *B. hinzii*, *B. pseudohinzii*, *B. holmesii*, *B. avium*, *B. trematum*, *B. ansorpii*, *B. flabilis*, *B. bronchealis*, *B. sputigena* (isolated from samples of human/animal origin), *B. petrii*, *B. tumbae*, *B. muralis*, and *B. tumulicola* (isolated from environmental samples). The corresponding strains were recovered from different environmental niches (Table 2), including soil (52 strains) and water (11 strains), and from 8 strains associated with plants. The soil samples were of diverse origin, including compost, cave rocks, and metal mines, but the majority were sampled at sites contaminated with oil and several halogenated cyclic hydrocarbons such as chlorinated benzenes or hexachlorocyclohexane. The samples from aquatic environments were also of diverse origin, namely industrial wastewater, a sulfur spring, lake water, surface sea water, and river biofilms. Several samples from plants were isolated from roots and thus at the plant-soil interface (Table 2). Thus, members of the genus *Bordetella* appear to be widespread across different environmental niches.

### 16S rRNA Gene Sequence Clades are Associated with Particular Environmental Niches

To relate the environmental isolates to known *Bordetella* species, we aligned the 16S rRNA gene sequences and constructed a Neighbor-joining tree using the Maximum-likelihood algorithm implemented in MEGA (Tamura et al., 2007). Forty-eight sequences from environmental samples were of sufficient length and used for further analyses (Table 2). Of those, 36 originated from soil (27 haplotypes), eight from aquatic environments (7 haplotypes), and four from plants (4 haplotypes). The tree was rooted with sequences of *Burkholderia pseudomallei* and *Ralstonia solanacearum* as outgroups. The *Bordetella* sequences formed 10 distinct clusters (Figure 1). While most clusters contained at least one described species, such as *B. petrii* in cluster VI or *B. tumbae/B. muralis* in cluster V, several *Bordetella* sequences did not cluster with any described species but rather occupied distinct branches of the tree. These include the two isolates in cluster IV, the isolates from soil samples in clusters VII and X and strains *B. sp.* CC-PW-55 and *B. sp.* TS-T34 (cluster

IX) isolated from surface sea water and lake water, respectively (Figure 1).

Superimposing the origin of the *Bordetella* spp. isolates revealed that most of the identified clusters were dominated by sequences of similar environmental/host origin. Thus, cluster I was composed of sequences of *B. holmesii* and the classical bordetellae (*B. bronchiseptica*, *B. parapertussis*, and *B. pertussis*), all of which were isolated from human and animal infection, but also contained *B. sp.* HT38 isolated from a river biofilm in China (Figure 1, Table 2). Cluster III contained sequences of species isolated from human respiratory specimen (*B. sputigena*, *B. bronchialis*, and *B. flabilis*) plus an isolate from soil in India. Other clusters either contained, or were dominated by, sequences of environmental origin such as cluster IV (water and soil), cluster V (soil), including the three species recovered from mural paintings *B. tumbae*, *B. tumulicola*, and *B. muralis*; but also *B. sp.* CCBAU from a maize rhizosphere and *B. ansorpii* from infection of an immunocompromised patient, and cluster VI (soil, including the environmental species *B. petrii*). The prominent exception to this pattern, cluster II, contained sequences from animal/human infection (*B. avium*, *B. hinzii*, *B. pseudohinzii*, and *B. trematum*) as well as from water (*B. sp.* MT-E1, *B. sp.* HF27), plant root (*B. sp.* R8-804, *B. sp.* R8-551), and soil samples (*B. sp.* BAB-4396). However, the other clusters were either dominated by animal-associated samples (clusters I and III) or samples of environmental origin (all other clusters).

If the genus *Bordetella* were of environmental origin, samples isolated from soil and water would be expected to be more diverse and would appear widespread across the tree. Indeed, environmental samples were present in all sequence clusters. In contrast, sequences from animal-associated samples were confined to four clusters, all of which also contained environmental isolates. Three of those four clusters formed a single super clade which originated from one of several clades among sequences from environmental isolates. In contrast, all clusters near the tree root exclusively contained environmental samples, but no animal associated samples (Figure 1). The phylogenetic analyses showed that the genetic diversity was significantly higher in sequences from environmental samples ( $\Pi_{95} = 2.02\text{--}2.13\%$ ) than in sequences from animal-associated samples ( $\Pi_{95} = 1.30\text{--}1.53\%$ ). The sequence of branching events within the phylogenetic tree is consistent with an environmental origin of *Bordetella* and subsequent adaptation of some lineages to animal hosts.

### *Bordetella bronchiseptica* and *Bordetella hinzii* are Capable of Growing in Soil Extract

Since most environmental *Bordetella* samples were recovered from soil (and water), we hypothesized that pathogenic, animal-associated species may have retained the ability to thrive in soil as an environmental niche. Therefore, we assessed the ability of *B. bronchiseptica* strain RB50, *B. hinzii* strain L60, and *B. petrii* strain DSMZ12804, to grow in a sterile, homogenized suspension made from soil. Instead of growing pathogenic bordetellae directly on solid soil, we prepared a soil suspension

TABLE 2 | *Bordetella* strains for which the 16S ribosomal RNA sequences were recovered from environmental samples.

<i>Bordetella</i> strains	Isolation source	Country	References	GenBank accession No.	Sequence length (bp)	Duplicated sequences
<b>SOIL ORIGIN</b>						
<i>Bordetella</i> sp. F2	Chlorinated benzenes polluted soil	Germany	Wang et al., 2007	DQ453689.1	1527	4
<i>Bordetella</i> sp. E3	Chlorinated benzenes polluted soil	Germany	Wang et al., 2007	DQ453688.1	1527	4
<i>Bordetella</i> sp. QJ2-5	Chlorinated benzenes polluted soil	China	NCBI	DQ152013.1	1393	4
<i>Bordetella</i> sp. 2b05	HCH-contaminated soil	India	NCBI	JF979304.1	1523	
<i>Bordetella</i> sp. 2f06	HCH-contaminated soil	India	NCBI	JF979347.1	1523	
<i>Bordetella</i> sp. 2e11	HCH-contaminated soil	India	NCBI	JF979341.1	1522	
<i>Bordetella</i> sp. 1h08	HCH-contaminated soil	India	NCBI	JF979288.1	1519	
<i>Bordetella</i> sp. 1c11	HCH-contaminated soil	India	NCBI	JF979241.1	1521	
<i>Bordetella</i> sp. 2c11	HCH-contaminated soil	India	NCBI	JF979320.1	1523	
<i>Bordetella</i> sp. 2a09	HCH-contaminated soil	India	NCBI	JF979298.1	1519	
<i>Bordetella</i> sp. ud1	1,2,4-TCB contaminated soil	Germany	NCBI	FJ529833.1	1523	4
<i>Bordetella</i> sp. ud29	1,2,4-TCB contaminated soil	Germany	NCBI	FJ529848.1	1523	
<i>Bordetella</i> sp. ud3b	1,2,4-TCB contaminated soil	Germany	NCBI	FJ529835.1	1523	4
<i>Bordetella</i> sp. ud13a	1,2,4-TCB contaminated soil	Germany	NCBI	FJ529840.1	1525	
<i>Bordetella</i> sp. IITR02	1,2,4-TCB contaminated soil	India	NCBI	EU752498.1	1422	
<i>Bordetella</i> sp. CTN-10	Chemical factory soil	China	NCBI	FJ598334.1	1398	2
<i>Bordetella</i> sp. 2-12	Chemical factory soil	China	NCBI	FJ598328.1	1410	2
<i>Bordetella</i> sp. CTN-16	Chemical factory soil	China	NCBI	FJ598326.1	1412	4
<i>Bordetella</i> sp. C16-Siri112	Oil-contaminated soil	Iran	NCBI	JX500276.1	1397	5
<i>Bordetella</i> sp. p23(2011)	Magnetite drainage, Iron mine	China	NCBI	HQ652588.1	1518	
<i>Bordetella</i> sp. e3(2011)	Magnetite drainage, Iron mine	China	NCBI	HQ652587.1	1501	3
<i>Bordetella</i> sp. d16(2011)	Magnetite drainage, Iron mine	China	NCBI	HQ652589.1	1507	3
<i>Bordetella</i> sp. f17(2011)	Magnetite drainage, Iron mine	China	NCBI	HQ652590.1	1520	
<i>Bordetella</i> sp. FB-8	Creek sediment from former uranium-mining area	Germany	NCBI	JN885794.1	1385	
<i>Bordetella</i> sp. A2-436	Uranium mine	Portugal	NCBI	KF441609.1	1528	
<i>Bordetella</i> sp. J4	Acid mine drainage	France	Delavat et al., 2013	HF568988.1	1410	
<i>Bordetella</i> sp. BAB-4396	Soil	India	NCBI	KM289182.1	1499	
<i>Bordetella</i> sp. B4	Paddy field by yellow river	China	NCBI	EU140499.1	1523	
<i>Bordetella</i> sp. MCYF11	Lake Taihu sediment	China	Yang et al., 2014	KC734882.1	1385	
<i>Bordetella</i> sp. PTG4-17	Sediment of the Indian ocean	India	NCBI	EU603444.1	1496	
<i>Bordetella</i> sp. RCC3	Caves rock	India	NCBI	KC119149.1	1476	
<i>Bordetella</i> sp. RCC4	Caves rock	India	NCBI	KC119150.1	1464	
<i>Bordetella</i> sp. M1-6	Compost	China	Kato et al., 2004	AB039335.1	1531	
<i>Bordetella</i> sp. FS1413	Compost	Finland	Partanen et al., 2010	FN667145.1	1464	
<i>Bordetella</i> sp. SMG22	Compost	China	Guo et al., 2015	AM930282.1	1491	6
<i>Bordetella</i> sp. OT-2-E7	Compost	China	Tian et al., 2013	JQ337611.1	1397	6
<i>Bordetella</i> sp. strain 2ABA4	Solid waste dumpsites	Nigeria	Sanuth et al., 2013	HE858274.1	1168	
<i>Bordetella</i> sp. OS17	Benten-Cho station soil	Japan	Matsumura et al., 2009	AB453298.1	980	
<i>Bordetella</i> sp. VVAR	Soil	Japan	NCBI	FJ588707.1	1451	
<i>Bordetella</i> sp. Ds-4	Cultivated soil	India	NCBI	HQ857791.1	727	
<i>Bordetella</i> sp. R-8	Garden soil	India	NCBI	JX130378.1	1319	
<i>Bordetella</i> sp. SPB-24	Garden soil	India	Bachate et al., 2012	JN208922.1	1403	
<i>Bordetella</i> sp. As3-3	Arsenic contaminated soil	Italy	Cavalca et al., 2010	FN392624.2	544	
<i>Bordetella</i> sp. AGO-03	Arsenic contaminated rice field	India	NCBI	AB696982.1	979	
<i>Bordetella</i> sp. ADP-18	Arsenic contaminated rice fields	India	NCBI	AB697485.1	674	
<i>Bordetella</i> sp. C16-Siri108	Oil-contaminated soil	Iran	NCBI	JX500272.1	1069	
<i>Bordetella</i> sp. C16-Siri113	Oil-contaminated soil	Iran	NCBI	JX500277.1	1295	
<i>Bordetella</i> sp. BF07B02	Agricultural soil	Burkina Faso	Colinon et al., 2013	KC195878.1	1381	

(Continued)



TABLE 2 | Continued

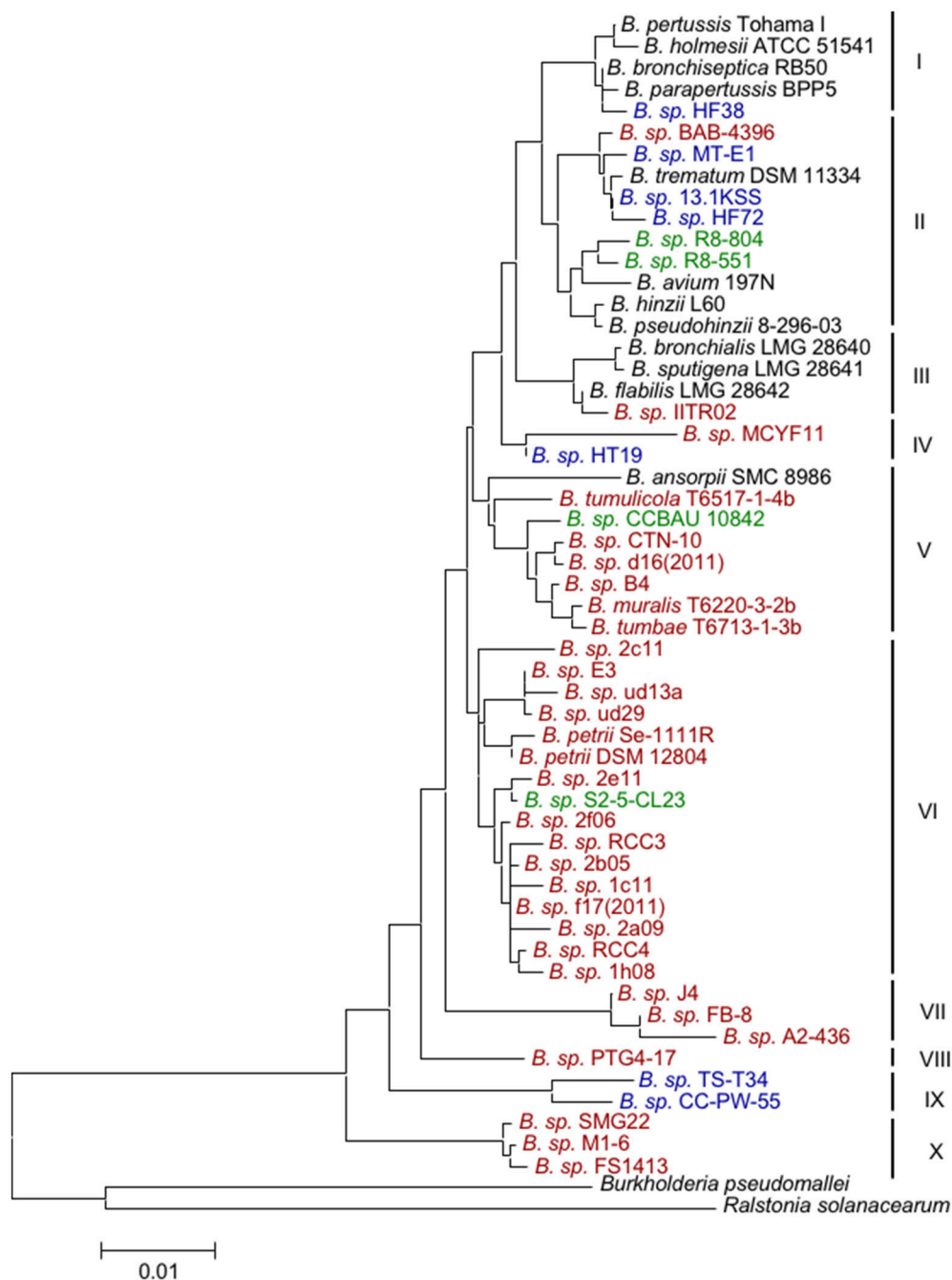
<i>Bordetella</i> strains	Isolation source	Country	References	GenBank accession No.	Sequence length (bp)	Duplicated sequences
<i>Bordetella</i> sp. HPC772	Activated sludge of an effluent treatment plant	India	NCBI	AY838357.1	580	
<i>Bordetella</i> sp. PH21	Phenolic compounds-contaminated sediment	China	NCBI	JN171686.1	721	
<i>Bordetella</i> sp. PH22	Phenolic compounds-contaminated sediment	China	NCBI	JN171687.1	721	
<i>Bordetella</i> sp. VKRKCd3	Seashore surface sediment	India	NCBI	GQ262759.1	363	
<b>PLANT ORIGIN</b>						
<b><i>Bordetella</i> sp. CCBau 10842</b>	Maize rhizosphere	China	NCBI	JF772555.1	1369	
<b><i>Bordetella</i> sp. R8-804</b>	<i>Jatropha curcas</i> L, plant root	Singapore	NCBI	JQ659985.1	1487	
<b><i>Bordetella</i> sp. R8-551</b>	<i>Jatropha curcas</i> L, plant root	Singapore	NCBI	JQ659951.1	1486	
<b><i>Bordetella</i> sp. S2-5-CL23</b>	velvetleaf seed	USA	NCBI	EU769148.1	1492	
<i>Bordetella</i> sp. S318(2010)	<i>M. sinensis</i> × <i>giganteus</i> internal stem tissue	Ireland	NCBI	HM102497.1	600	
<i>Bordetella</i> sp. Juv992	Lupine cluster roots	Switzerland	Weisskopf et al., 2011	JN590346.1	1302	
<i>Bordetella</i> sp. PnB 4	Pepper	India	NCBI	JQ886795.1	370	
<i>Bordetella</i> sp. RS-CIW-47	Maize rhizosphere	Pakistan	NCBI	KC430988.1	950	
<b>WATER ORIGIN</b>						
<b><i>Bordetella</i> sp. MT-I2</b>	Industrial wastewater	Germany	Toups et al., 2010	EU727195.1	1526	1
<b><i>Bordetella</i> sp. MT-E1</b>	Industrial wastewater	Germany	Toups et al., 2010	EU727194.1	1525	1
<b><i>Bordetella</i> sp. TS-T34</b>	Lake water	China	NCBI	KC762319.1	1398	
<b><i>Bordetella</i> sp. CC-PW-55</b>	Surface seawater	Taiwan	NCBI	KF851340.1	1500	
<b><i>Bordetella</i> sp. 13.1 KSS</b>	Mineral oil-based metalworking fluid	Germany	Lodders and Kämpfer, 2012	HE575910.1	1398	
<b><i>Bordetella</i> sp. HT19</b>	Sulfur spring	India	NCBI	FJ969843.1	1404	
<b><i>Bordetella</i> sp. HF38</b>	River biofilms	China	NCBI	KR188914.1	1523	
<b><i>Bordetella</i> sp. HF72</b>	River biofilms	China	NCBI	KR188948.1	1523	
<i>Bordetella</i> sp. MMJ09	Distillery wastewater	China	NCBI	GU244378.1	813	
<i>Bordetella</i> sp. Sulf-8	Municipal wastewater	South Korea	Nisola et al., 2010	GU812430.1	1314	
<i>Bordetella</i> sp. IPJ1	Rusted iron pipe in freshwater lake	India	NCBI	HM593901.1	1100	

In bold are the strains for which the length of the 16S ribosomal RNA sequence were at least 1376 bp, and were included in the phylogenetic tree construction.

to extract possible nutrients but to avoid solid matter which allowed visual monitoring of bacterial growth and selection of appropriate sampling time points. All three isolates were cultured at room temperature (25°C) with shaking in either liquid soil extract or in Stainer-Scholte (SS) medium as a control. All three species grew fast in SS medium with doubling times of  $1.8 \pm 0.02$  h (*B. bronchiseptica*),  $1.9 \pm 0.01$  h (*B. hinzii*), and  $1.9 \pm 0.02$  h (*B. petrii*), and reached the stationary phase prior to 48 h post-inoculation (Figure 2). As expected from an environmental bacterium, *B. petrii* strain DSMZ12804 thrived when inoculated into a soil extract, with a doubling time of  $7.25 \pm 0.24$  h (Figure 2). Surprisingly, both *B. hinzii* strain L60 with a doubling time of  $6.4 \pm 0.09$  h and *B. bronchiseptica* strain RB50 with a doubling time of  $4.0 \pm 0.04$  h grew in the soil extract faster than *B. petrii*. Thus, all three species can grow efficiently at 25°C on filter-sterilized soil extract, even though the growth rate was slower than in Stainer-Scholte medium.

## DISCUSSION

Bacteria of the genus *Bordetella* occupy remarkably diverse ecological niches, ranging from soil, water, and plants, to the respiratory tracts of a wide variety of animals including humans. Several environmental *Bordetella* strains were isolated from soils polluted with oil and oil derivatives (Table 2), including halogenated polycyclic hydrocarbons (Eriksson et al., 2003; Bianchi et al., 2005; Wang et al., 2007). Other strains were found in garden soil, compost, and various sediments suggesting these organisms are quite adaptable to diverse sites. The only sequenced and analyzed genome of an environmental isolate, *B. petrii* strain DSMZ 12804, revealed a possible genomic basis for substantial metabolic versatility (Gross et al., 2008). The genome encodes multiple auxiliary pathways for the utilization of a variety of nutrients, including pectate, numerous sugar derivatives from degraded plant products and various aromatic

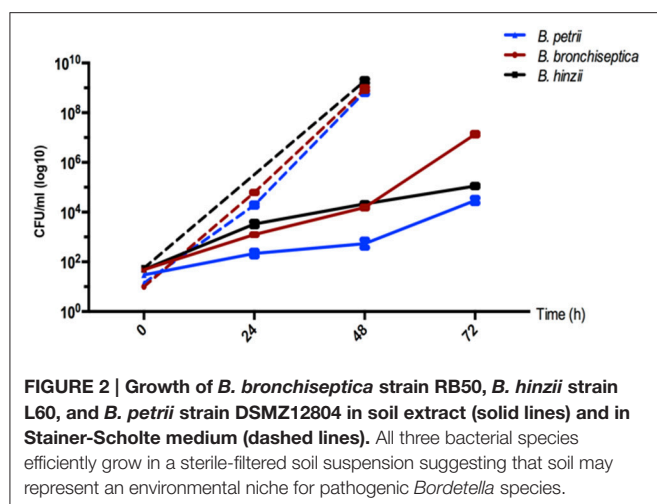


**FIGURE 1 | Neighbor-Joining tree based on 16S rRNA gene sequences of animal-associated and environmental strains of *Bordetella*.** The 52 near full-length sequences (1376 bp) formed 10 clades (I–X) of phylogenetically closely related *Bordetella* isolates/species recovered from soil (brown), water (blue), plants (green) and animals (black). The 16S rRNA gene sequences of the beta-proteobacteria *Burkholderia pseudomallei* and *Ralstonia solanacearum* were used as outgroups.

compounds. Five of the eight genomic islands that have been identified in this genome contain genes coding for enzymes for the metabolism of aromatic compounds, particularly clusters of genes encoding enzymes of the chlorocatechol pathway, including gene clusters that show high similarity to genes in a 1,2,4-trichlorobenzene-degrading *Pseudomonas* strain (Gross

et al., 2008). The presence of multiple chlorocatechol gene clusters in addition to several different central pathways for aromatic metabolism may provide a competitive advantage for growth in contaminated environments.

Another striking feature of environmental *Bordetella* isolates is their resistance to heavy metals (Cavalca et al., 2010). Ten out



of 52 soil samples (Table 2) were isolated from iron mines (e.g., *B. sp.* d16, *B. sp.* f17), from uranium mines (*B. sp.* FB-8, *B. sp.* A2-436), or from soil polluted with arsenic (e.g., *B. sp.* As3-3). Such remarkable metal tolerance is most likely conferred by heavy metal resistance systems. Indeed, the genome of *B. petrii* strain DSMZ 12804 contains several heavy metal resistance operons on a genomic island absent from the genomes of other sequenced bordetellae, whereas other strains contain different islands of genes. Ultimately, the presence of multiple heavy metal resistance systems may allow environmental *Bordetella* isolates to thrive in metal rich environments.

Most plant-associated *Bordetella* strains were recovered from roots (*B. sp.* R8-804, *B. sp.* R8-551, *B. sp.* Juv992) and the rhizosphere at the plant-soil interface (*B. sp.* CCBAU 10842). Thus, these isolates may in fact represent soil samples or, alternatively, may be involved in interactions with plants at the plant-soil interface. The resemblance between plant responses to bacterial virulence factors and the responses of mammalian immune cells (Berg et al., 2005) serve as evidence that bacteria-plant interactions may have paved the way for bacterial adaptation to animals. In this regard, plant-root isolates *B. sp.* R8-804 and *B. sp.* R8-551 from plant roots are closely related to bird pathogens, *B. hinzii* and *B. avium*, supporting the view that plants cells could serve as a “training ground” for environmental strains that eventually gain the ability to colonize animal hosts (Berg et al., 2005).

In addition to these plant root isolates, several other environmental isolates were also found to be very closely related to animal-associated pathogens (Figure 1). Interestingly, those strains were isolated from very diverse sources, namely (polluted) soil in India (*B. sp.* BAB-4396, *B. sp.* IITR02), industrial waste water (*B. sp.* MT-E1), and oil-based metal-working emulsion in Germany (*B. sp.* 13.1 KSS), as well as from river biofilms in China (*B. sp.* HF38 and *B. sp.* HF72). The two isolates from a river biofilm in China are of particular interest. The 16S rRNA sequence of one of those (strain HF72) showed 99.56% sequence similarity to that of the human pathogen *B. trematum* (6 SNPs). According to 16S rRNA gene sequence, the other isolate (*B. sp.* HF38) is even more closely related (99.78%, three SNPs) to the

animal pathogen *B. bronchiseptica* strain RB50 and the human pathogen *B. parapertussis* strain 12822, which share an identical sequence in this gene. By this measure, isolate HF38 is as closely related to *B. bronchiseptica* strain RB50 and *B. parapertussis* strain 12822 as it is to *B. pertussis*. This exceptionally close phylogenetic relatedness makes several evolutionary scenarios conceivable. First, isolate *B. sp.* HF38 may be an environmental, non-pathogenic strain closely related to the animal/human pathogens among the classical bordetellae. Second, this isolate might be a descendant or relative of an ancestor of the classical bordetellae which later became pathogenic after acquisition of several virulence-associated factors, such as pertussis toxin, adenylate cyclase toxin, and dermonecrotic toxin. Third, this isolate may in fact represent a *B. bronchiseptica* or *B. parapertussis* strain that naturally survives and/or grows within an environmental reservoir. Although the classical bordetellae have not yet been isolated from outside a mammalian host, our results suggest that animal-pathogenic *Bordetella* species retain the ability to grow in soil as an environmental niche. This implies that *B. bronchiseptica* and other species might be found (at least transiently) in soil, for example at farms with suitable animal hosts such as cattle, pig, sheep and horse, or near dog kennels. Interestingly, even fastidious *B. pertussis* bacteria remained able to be cultured for up to 5 days when spread onto various hospital-setting surfaces such as fabrics, plastics, glass, and paper, and also in several infant foods (Ocklitz and Milleck, 1967). Fourth, *B. sp.* HF38 as well as other isolates from water and soil may be protected internally by a non-vertebrate host. For example, amoebae are known to host bacteria such as *Legionella pneumophila* (Molmeret et al., 2005), and amoeba-grown *L. pneumophila* exhibited radically increased resistance to harsh environmental conditions such as fluctuations in temperature, osmolarity, acidity, as well as to biocides that may facilitate bacterial survival and persistence in the environment (Barker et al., 1995; Abu Kwaik et al., 1997, 1998; Winiecka-Krusnell and Linder, 1999). Amoebae are ubiquitously found in most environments, and shared habitats between amoeba and *Bordetella* could be an important factor for the persistence of the bacteria. Indeed, our group has shown that the animal-adapted *B. bronchiseptica* is able to survive and multiply intracellularly in the trophozoites and sori of the amoeba *Dictyostelium discoideum* before being disseminated with the amoeba spores to novel geographical locations (Bendor et al., in revision). Thus, in addition to our recent data demonstrating that *B. bronchiseptica* can circulate and efficiently transmit amongst mammals, these data demonstrate that this species can also grow and disseminate efficiently in association with amoebae. These independent but interconnected *Bordetella* lifecycles allow for disease propagation, transmission, and re-emergence in the absence of an infected animal host.

Strains included in this study were identified as *Bordetella* spp. based on their 16S rRNA gene sequence. Currently, there are no data available regarding potential pathogenicity of these species. Whole genome sequencing will provide valuable insights into the evolution and ecology of environmental vs. animal-pathogenic bordetellae. Of special interest are environmental isolates closely related to animal pathogens, particularly isolate *B. sp.* HF38, and analysis of their genomes will reveal whether they are

non-pathogenic relatives of known animal pathogens or if they in fact represent environmental reservoirs of *B. bronchiseptica* or *B. parapertussis*.

Finally, the majority of environmental *B. sp.* were recovered from soil samples indicating that soil could be the most frequent natural habitat of bordetellae. Indeed, sequences identified from soil samples were found in 8 of 10 sequence clusters, including samples from compost in cluster X at the root of the tree (**Figure 1**). The sequence of branching events within the phylogenetic tree, the significantly higher sequence diversity in samples from soil and water than in those from animals, as well as the preserved ability of animal pathogens to grow in soil, suggest an environmental, likely soil-based, origin of the genus *Bordetella*. Thus, similar to bacteria of the closely related genus *Achromobacter*, which are of environmental origin but also contain opportunistic pathogens (Li et al., 2013), *Bordetella*

appears to be a bacterium of environmental origin that adapted and became pathogenic via the acquisition of factors mediating specific interactions with animal hosts.

## AUTHOR CONTRIBUTIONS

IHS, BL, and ETH conceived and designed the experiments. IHS and BL performed the experiments and analyzed the data. IHS, BL and ETH wrote the paper.

## ACKNOWLEDGMENTS

We thank Holly Vuong, Monica Cartelle Gestal, and Israel Rivera from the Harvill lab for helpful discussions. This work was supported by grants GM113681 and AI116186 by the National Institutes of Health (to ETH).

## REFERENCES

- Abu Kwaik, Y., Gao, L. Y., Harb, O. S., and Stone, B. J. (1997). Transcriptional regulation of the macrophage-induced gene (*gspA*) of *Legionella pneumophila* and phenotypic characterization of a null mutant. *Mol. Microbiol.* 24, 629–642. doi: 10.1046/j.1365-2958.1997.3661739.x
- Abu Kwaik, Y., Gao, L. Y., Stone, B. J., Venkataraman, C., and Harb, O. S. (1998). Invasion of protozoa by *Legionella pneumophila* and its role in bacterial ecology and pathogenesis. *Appl. Environ. Microbiol.* 64, 3127–3133.
- Bachate, S. P., Khapare, R. M., and Kodam, K. M. (2012). Oxidation of arsenite by two  $\beta$ -proteobacteria isolated from soil. *Appl. Microbiol. Biotechnol.* 93, 2135–2145. doi: 10.1007/s00253-011-3606-7
- Barker, J., Scaife, H., and Brown, M. R. (1995). Intraphagocytic growth induces an antibiotic-resistant phenotype of *Legionella pneumophila*. *Antimicrob. Agents Chemother.* 39, 2684–2688. doi: 10.1128/AAC.39.12.2684
- Berg, G., Eberl, L., and Hartmann, A. (2005). The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ. Microbiol.* 7, 1673–1685. doi: 10.1111/j.1462-2920.2005.00891.x
- Bianchi, F., Careri, M., Mustat, L., Malcevski, A., and Musci, M. (2005). Bioremediation of toluene and naphthalene: development and validation of a GC-FID method for their monitoring. *Ann. Chim.* 95, 515–524. doi: 10.1002/adich.200590061
- Biederman, L., Rosen, M. R., Bobik, B. S., and Roberts, A. L. (2015). *Bordetella pertussis* recovered from chronic pansinusitis in an adult with cystic fibrosis. *IDCases* 2, 97–98. doi: 10.1016/j.idcr.2015.09.004
- Cavalca, L., Zanchi, R., Corsini, A., Colombo, M., Romagnoli, C., Canzi, E., et al. (2010). Arsenic-resistant bacteria associated with roots of the wild *Cirsium arvense* (L.) plant from an arsenic polluted soil, and screening of potential plant growth-promoting characteristics. *Syst. Appl. Microbiol.* 33, 154–164. doi: 10.1016/j.syapm.2010.02.004
- Colinon, C., Deredjian, A., Hien, E., Brothier, E., Bouziri, L., Cournoyer, B., et al. (2013). Detection and enumeration of *Pseudomonas aeruginosa* in soil and manure assessed by an ecfX qPCR assay. *J. Appl. Microbiol.* 114, 1734–1749. doi: 10.1111/jam.12189
- Delavat, F., Lett, M. C., and Lièvremon, D. (2013). Yeast and bacterial diversity along a transect in an acidic, As-Fe rich environment revealed by cultural approaches. *Sci. Total Environ.* 463–464, 823–828. doi: 10.1016/j.scitotenv.2013.06.023
- Diavatopoulos, D. A., Cummings, C. A., Schouls, L. M., Brinig, M. M., Relman, D. A., and Mooi, F. R. (2005). *Bordetella pertussis*, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of *B. bronchiseptica*. *PLoS Pathog.* 1:e45. doi: 10.1371/journal.ppat.0010045
- Eriksson, M., Sodersten, E., Yu, Z., Dalhammar, G., and Mohn, W. W. (2003). Degradation of polycyclic aromatic hydrocarbons at low temperature under aerobic and nitrate-reducing conditions in enrichment cultures from northern soils. *Appl. Environ. Microbiol.* 69, 275–284. doi: 10.1128/AEM.69.1.275-284.2003
- Fry, N. K., Duncan, J., Malnick, H., Warner, M., Smith, A. J., Jackson, M. S., et al. (2005). *Bordetella pertussis* clinical isolate. *Emerging Infect. Dis.* 11, 1131–1133. doi: 10.3201/eid1107.050046
- Goodnow, R. A. (1980). Biology of *Bordetella bronchiseptica*. *Microbiol. Rev.* 44, 722–738.
- Gross, R., Guzman, C. A., Sebahia, M., dos Santos, V. A., Pieper, D. H., Koebnik, R., et al. (2008). The missing link: *Bordetella pertussis* is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. *BMC Genomics* 9:449. doi: 10.1186/1471-2164-9-449
- Guo, Y., Zhang, J., Yan, Y., Wu, J., Zhu, N., and Deng, C. (2015). Molecular phylogenetic diversity and spatial distribution of bacterial communities in cooling stage during swine manure composting. *Asian-Australas. J. Anim. Sci.* 28, 888–895. doi: 10.5713/ajas.14.0882
- Ivanov, Y. V., Linz, B., Register, K. B., Newman, J. D., Taylor, D. L., Boschert, K. R., et al. (2016). Identification and taxonomic characterization of *Bordetella pseudohinzii* sp. nov. isolated from laboratory-raised mice. *Int. J. Syst. Evol. Microbiol.* 66, 5452–5459. doi: 10.1099/ijsem.0.001540
- Ivanov, Y. V., Shariat, N., Register, K. B., Linz, B., Rivera, I., Hu, K., et al. (2015). A newly discovered *Bordetella* species carries a transcriptionally active CRISPR-Cas with a small Cas9 endonuclease. *BMC Genomics* 16:863. doi: 10.1186/s12864-015-2028-9
- Johnson, S. L., Bishop-Lilly, K. A., Ladner, J. T., Daligault, H. E., Davenport, K. W., Jaissle, J., et al. (2015). Complete genome sequences for 59 *Burkholderia* isolates, both pathogenic and near neighbor. *Genome Announc.* 3:e00159-15. doi: 10.1128/genomeA.00159-15
- Kato, S., Haruta, S., Cui, Z. J., Ishii, M., and Igarashi, Y. (2004). Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria. *FEMS Microbiol. Ecol.* 51, 133–142. doi: 10.1016/j.femsec.2004.07.015
- Kattar, M. M., Chavez, J. F., Limaye, A. P., Rassoul-Barrett, S. L., Yarfitz, S. L., Carlson, L. C., et al. (2000). Application of 16S rRNA gene sequencing to identify *Bordetella hinzii* as the causative agent of fatal septicemia. *J. Clin. Microbiol.* 38, 789–794.
- Kerstens, K., Hinz, K. H., Hertle, A., Segers, P., Lievens, A., Siegmund, O., et al. (1984). *Bordetella avium* sp. nov., isolated from the respiratory tracts of turkeys and other birds. *Int. J. Syst. Bacteriol.* 34, 56–70. doi: 10.1099/00207173-34-1-56
- Kigawa, R., Sano, C., Nishijima, M., Tazato, N., Kiyuna, T., Hayakawa, N., et al. (2013). Investigation of acetic acid bacteria isolated from the Kitora tumulus in Japan and their involvement in the deterioration of the plaster of the mural paintings. *Stud. Conserv.* 58, 30–40. doi: 10.1179/2047058412Y.0000000040
- Ko, K. S., Peck, K. R., Oh, W. S., Lee, N. Y., Lee, J. H., and Song, J. H. (2005). New species of *Bordetella*, *Bordetella ansorpii* sp. nov., isolated from



- the purulent exudate of an epidermal cyst. *J. Clin. Microbiol.* 43, 2516–2519. doi: 10.1128/JCM.43.5.2516-2519.2005
- Li, X., Hu, Y., Gong, J., Zhang, L., and Wang, G. (2013). Comparative genome characterization of *Achromobacter* members reveals potential genetic determinants facilitating the adaptation to a pathogenic lifestyle. *Appl. Microbiol. Biotechnol.* 97, 6413–6425. doi: 10.1007/s00253-013-5018-3
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Lodders, N., and Kämpfer, P. (2012). A combined cultivation and cultivation-independent approach shows high bacterial diversity in water-miscible metalworking fluids. *Syst. Appl. Microbiol.* 35, 246–252. doi: 10.1016/j.syapm.2012.03.006
- Matsumura, Y., Hosokawa, C., Sasaki-Mori, M., Akahira, A., Fukunaga, K., Ikeuchi, T., et al. (2009). Isolation and characterization of novel bisphenol-A-degrading bacteria from soils. *Biocontrol Sci.* 14, 161–169. doi: 10.4265/bio.14.161
- Mattoo, S., and Cherry, J. D. (2005). Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. *Clin. Microbiol. Rev.* 18, 326–382. doi: 10.1128/CMR.18.2.326-382.2005
- Molmeret, M., Horn, M., Wagner, M., Santic, M., and Abu Kwaik, Y. (2005). Amoebae as training grounds for intracellular bacterial pathogens. *Appl. Environ. Microbiol.* 71, 20–28. doi: 10.1128/AEM.71.1.20-28.2005
- Musser, J. M., Hewlett, E. L., Peppler, M. S., and Selander, R. K. (1986). Genetic diversity and relationships in populations of *Bordetella* spp. *J. Bacteriol.* 166, 230–237. doi: 10.1128/jb.166.1.230-237.1986
- Nagata, J. M., Charville, G. W., Klotz, J. M., Wickremasinghe, W. R., Kann, D. C., Schwenk, H. T., et al. (2015). *Bordetella petrii* sinusitis in an immunocompromised adolescent. *Pediatr. Infect. Dis. J.* 34, 458. doi: 10.1097/INF.0000000000000564
- Nisola, G. M., Tuuguu, E., Farnazo, D. M., Han, M., Kim, Y., Cho, E., et al. (2010). Hydrogen sulfide degradation characteristics of *Bordetella* sp. Sulf-8 in a biotrickling filter. *Bioprocess. Biosyst. Eng.* 33, 1131–1138. doi: 10.1007/s00449-010-0440-8
- Ocklitz, H. W., and Milleck, J. (1967). Die überlebenszeit von pertussisbakterien außerhalb des kranken. experimentelle untersuchungen zur keuchhustenenepidemiologie. *Zentralblatt Bakteriologie Parasitenkunde Infektionskrankheiten* 203, 79–91.
- Park, J., Zhang, Y., Buboltz, A. M., Zhang, X., Schuster, S. C., Ahuja, U. et al. (2012). Comparative genomics of the classical *Bordetella* subspecies: the evolution and exchange of virulence-associated diversity amongst closely related pathogens. *BMC Genomics* 13:545. doi: 10.1186/1471-2164-13-545
- Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., et al. (2003). Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* 35, 32–40. doi: 10.1038/ng1227
- Partanen, P., Hultman, P., Paulin, L., Auvinen, P., and Romantschuk, M. (2010). Bacterial diversity at different stages of the composting process. *BMC Microbiol.* 10:94. doi: 10.1186/1471-2180-10-94
- Register, K. B., Ivanov, Y. V., Jacobs, N., Meyer, J. A., Goodfield, L. L., Muse, S. J. et al. (2015). Draft genome sequences of 53 genetically distinct isolates of *Bordetella bronchiseptica* representing 11 terrestrial and aquatic hosts. *Genome Announc.* 3:e00152-15. doi: 10.1128/genomeA.00152-15
- Register, K. B., and Kunkle, R. A. (2009). Strain-specific virulence of *Bordetella hinzii* in poultry. *Avian Dis.* 53, 50–54. doi: 10.1637/8388-070108-Reg.1
- Sanuth, H. A., Yadav, A., Fagade, O. E., and Shouche, Y. (2013). epsilon-caprolactam utilization by *Proteus* sp. and *Bordetella* sp. Isolated from solid waste dumpsites in Lagos State, Nigeria, first report. *Indian J. Microbiol.* 53, 221–226. doi: 10.1007/s12088-013-0356-5
- Sebaihia, M., Preston, A., Maskell, D. J., Kuzmiak, H., Connell, T. D., King, N. D. et al. (2006). Comparison of the genome sequence of the poultry pathogen *Bordetella avium* with those of *B. bronchiseptica*, *B. pertussis*, and *B. parapertussis* reveals extensive diversity in surface structures associated with host interaction. *J. Bacteriol.* 188, 6002–6015. doi: 10.1128/JB.01927-05
- Shepard, C. W., Daneshvar, M. I., Kaiser, R. M., Ashford, D. A., Lonsway, D., Patel, J. B., et al. (2004). *Bordetella holmesii* bacteremia: a newly recognized clinical entity among asplenic patients. *Clin. Infect. Dis.* 38, 799–804. doi: 10.1086/381888
- Stainer, D. W., and Scholte, M. J. (1970). A simple chemically defined medium for the production of phase I *Bordetella pertussis*. *J. Gen. Microbiol.* 63, 211–220. doi: 10.1099/00221287-63-2-211
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Tazato, N., Handa, Y., Nishijima, M., Kigawa, R., Sano, C., and Sugiyama, J. (2015). Novel environmental *Bordetella* species isolated from the plaster wall surface of mural paintings in the Takamatsuzuka tumulus: *Bordetella muralis* sp. nov., *Bordetella tumulicola* sp. nov. and *Bordetella tumbae* sp. nov. *Int. J. Syst. Evol. Microbiol.* 65, 4830–4838. doi: 10.1099/ijsem.0.000655
- Tian, W., Sun, Q., Xu, D. B., Zhang, Z. H., Chen, D., Li, C. Y., et al. (2013). Succession of bacterial communities during composting process as detected by 16S rRNA clone libraries analysis. *Int. Biodeterior. Biodegradation* 78, 58–66. doi: 10.1016/j.ibiod.2012.12.008
- Toups, M., Wübbeler, J. H., and Steinbüchel, A. (2010). Microbial utilization of the industrial wastewater pollutants 2-ethylhexylthioglycolic acid and iso-octylthioglycolic acid by aerobic gram-negative bacteria. *Biodegradation* 21, 309–319. doi: 10.1007/s10532-009-9302-y
- Vandamme, P. A., Peeters, C., Cnockaert, M., Inganäs, E., Falsen, E., Moore, E. R., et al. (2015). *Bordetella bronchialis* sp. nov., *Bordetella flabilis* sp. nov. and *Bordetella sputigena* sp. nov., isolated from human respiratory specimens, and reclassification of *Achromobacter sediminum* Zhang et al. 2014 as *Verticium sediminum* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 65, 3674–3682. doi: 10.1099/ijsem.0.000473
- Vandamme, P., Heyndrickx, M., Vancanneyt, M., Hoste, B., De Vos, P., Falsen, E., et al. (1996). *Bordetella trematum* sp. nov., isolated from wounds and ear infections in humans, and reassessment of *Alcaligenes denitrificans* Ruger and Tan 1983. *Int. J. Syst. Bacteriol.* 46, 849–858. doi: 10.1099/00207713-46-4-849
- Vandamme, P., Hommez, J., Vancanneyt, M., Monsieurs, M., Hoste, B., Cookson, B., et al. (1995). *Bordetella hinzii* sp. nov., isolated from poultry and humans. *Int. J. Syst. Bacteriol.* 45, 37–45. doi: 10.1099/00207713-45-1-37
- von Wintzingerode, F., Schattke, A., Siddiqui, R. A., Rösick, U., Göbel, U. B., and Gross, R. (2001). *Bordetella petrii* sp. nov., isolated from an anaerobic bioreactor, and emended description of the genus *Bordetella*. *Int. J. Syst. Evol. Microbiol.* 51, 1257–1265. doi: 10.1099/00207713-51-4-1257
- Wang, F., Grundmann, S., Schmid, M., Dörfler, U., Rohrer, S., Charles Munch, J., et al. (2007). Isolation and characterization of 1,2,4-trichlorobenzene mineralizing *Bordetella* sp. and its bioremediation potential in soil. *Chemosphere* 67, 896–902. doi: 10.1016/j.chemosphere.2006.11.019
- Weisskopf, L., Heller, S., and Eberl, L. (2011). *Burkholderia* species are major inhabitants of white lupin cluster roots. *Appl. Environ. Microbiol.* 77, 7715–7720. doi: 10.1128/AEM.05845-11
- Weyant, R. S., Hollis, D. G., Weaver, R. E., Amin, M. F., Steigerwalt, A. G., O'Connor, S. P., et al. (1995). *Bordetella holmesii* sp. nov., a new gram-negative species associated with septicemia. *J. Clin. Microbiol.* 33, 1–7.
- Winiecka-Krusnell, J., and Linder, E. (1999). Free-living amoebae protecting *Legionella* in water: the tip of an iceberg? *Scand. J. Infect. Dis.* 31, 383–385. doi: 10.1080/00365549950163833
- Yang, F., Zhou, Y., Yin, L., Zhu, G., Liang, G., and Pu, Y. (2014). Microcystin-degrading activity of an indigenous bacterial strain *Stenotrophomonas acidaminiphila* MC-LTH2 isolated from Lake Taihu. *PLoS ONE* 9:e86216. doi: 10.1371/journal.pone.0086216

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hamidou Soumana, Linz and Harvill. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Marco Rinaldo Oggioni,  
University of Leicester,  
United Kingdom  
Giovanni Delogu,  
Università Cattolica del Sacro Cuore,  
Italy  
Andrej Trauner,  
Swiss Tropical and Public Health  
Institute, Switzerland

### \*Correspondence:

Laura Pérez-Lago  
lperezg00@gmail.com  
Dario García-de-Viedma  
dgviedma2@gmail.com

†These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 13 October 2017

**Accepted:** 20 December 2017

**Published:** 19 January 2018

### Citation:

Herranz M, Pole I, Ozere I,  
Chiner-Oms Á, Martínez-Lirola M,  
Pérez-García F, Gijón P, Serrano MJR,  
Romero LC, Cuevas O, Comas I,  
Bouza E, Pérez-Lago L and  
García-de-Viedma D (2018)  
*Mycobacterium tuberculosis* Acquires  
Limited Genetic Diversity in Prolonged  
Infections, Reactivations and  
Transmissions Involving Multiple  
Hosts. *Front. Microbiol.* 8:2661.  
doi: 10.3389/fmicb.2017.02661

# *Mycobacterium tuberculosis* Acquires Limited Genetic Diversity in Prolonged Infections, Reactivations and Transmissions Involving Multiple Hosts

Marta Herranz<sup>1,2,3</sup>, Ilva Pole<sup>4,5</sup>, Iveta Ozere<sup>4,6</sup>, Álvaro Chiner-Oms<sup>7</sup>,  
Miguel Martínez-Lirola<sup>8</sup>, Felipe Pérez-García<sup>1,2</sup>, Paloma Gijón<sup>1,2,3</sup>,  
María Jesús Ruiz Serrano<sup>1,2,3</sup>, Laura Clotet Romero<sup>9</sup>, Oscar Cuevas<sup>10</sup>, Iñaki Comas<sup>11,12</sup>,  
Emilio Bouza<sup>1,2,3,13</sup>, Laura Pérez-Lago<sup>1,2,3\*†</sup> and Dario García-de-Viedma<sup>1,2,3\*†</sup>

<sup>1</sup> Servicio Microbiología Clínica y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón, Madrid, Spain,

<sup>2</sup> Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain, <sup>3</sup> CIBER Enfermedades Respiratorias (CIBERES),  
Madrid, Spain, <sup>4</sup> Childhood Tuberculosis Department, Centre of Tuberculosis and Lung Diseases, Riga East University  
Hospital, Riga, Latvia, <sup>5</sup> Latvian Biomedical Research and Study Centre, Riga, Latvia, <sup>6</sup> Department of Infectology and  
Dermatology, Riga Stradiņš University, Riga, Latvia, <sup>7</sup> Unidad Mixta Genómica y Salud, Centro Superior de Investigación en  
Salud Pública (FISABIO)-Universitat de València, Valencia, Spain, <sup>8</sup> Servicio de Microbiología, Complejo Hospitalario  
Torrecárdenas, Almería, Spain, <sup>9</sup> Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública al Vallès  
Occidental i Vallès Oriental, Subdirecció General de Vigilància i Resposta a Emergències de Salut Pública, Agència de Salut  
Pública de Catalunya, Barcelona, Spain, <sup>10</sup> Servicio de Laboratorio, Institut d'Investigació i Innovació Parc Taulí, I3PT Parc  
Taulí Hospital Universitari, Universitat Autònoma de Barcelona, Barcelona, Spain, <sup>11</sup> Instituto de Biomedicina de Valencia,  
Consejo Superior de Investigaciones Científicas, Valencia, Spain, <sup>12</sup> CIBER en Epidemiología y Salud Pública, Madrid, Spain,  
<sup>13</sup> Departamento de Medicina, Facultad de Medicina, Universidad Complutense de Madrid, Madrid, Spain

**Background:** *Mycobacterium tuberculosis* (MTB) has limited ability to acquire variability. Analysis of its microevolution might help us to evaluate the pathways followed to acquire greater infective success. Whole-genome sequencing (WGS) in the analysis of the transmission of MTB has elucidated the magnitude of variability in MTB. Analysis of transmission currently depends on the identification of clusters, according to the threshold of variability (<5 SNPs) between isolates.

**Objective:** We evaluated whether the acquisition of variability in MTB, was more frequent in situations which could favor it, namely inpatient, prolonged infections or reactivations and interpatient transmissions involving multiple sequential hosts.

**Methods:** We used WGS to analyze the accumulation of variability in sequential isolates from prolonged infections or translations from latency to reactivation. We then measured microevolution in transmission clusters with prolonged transmission time, high number of involved cases, simultaneous involvement of latency and active transmission.

**Results:** Inpatient and interpatient acquisition of variability was limited, within the ranges expected according to the thresholds of variability proposed, even though bursts of variability were observed.

**Conclusions:** The thresholds of variability proposed for MTB seem to be valid in most circumstances, including those theoretically favoring acquisition of variability. Our data point to multifactorial modulation of microevolution, although further studies are necessary to elucidate the factors underlying this modulation.

**Keywords:** tuberculosis, variability, whole genome sequencing, SNPs, microevolution

## INTRODUCTION

Whole-genome sequencing (WGS) has transformed the way we analyze transmission of tuberculosis (TB) (Nikolayevskyy et al., 2016; Comas, 2017; Satta et al., 2017). Identification of transmission clusters is currently based on determination of the magnitude of genomic diversity among the isolates in a population. An exhaustive analysis of the magnitude of this variability acquired in different clinical/epidemiological situations led to the definition of thresholds to determine whether 2 isolates were part of the same transmission chain, i.e., clustered (<5 SNPs) or unrelated (>12 SNPs) (Walker et al., 2013). Since then, these thresholds have been used as a consensus reference, and it has been accepted that microevolution in *Mycobacterium tuberculosis* (MTB) infection occurs within those ranges and with no wider deviations expected (Casali et al., 2016; Seto et al., 2017; Witney et al., 2017).

Acquisition of variability in MTB has been studied not only to determine the SNP thresholds to be applied in genomic epidemiology, but also to help us to understand one of the microevolution paths in MTB infection which may lead to more successful variants, in addition to other molecular events such as insertion/deletions or intragenomic recombination.

Different studies have addressed the emergence of clonal variants in MTB infections according to different molecular markers (Warren et al., 2002; Shamputa et al., 2006; Al-Hajj et al., 2010; Navarro et al., 2011; Black et al., 2015). The magnitude of inpatient diversity which can be measured for a strain within the same patient has been found to be comparable to the one observed after sequential rounds of multiple hosts sequentially along transmission chains (Pérez-Lago et al., 2014). Sometimes, these variants have been found associated to prolonged infections or to a history of previous TB (Navarro et al., 2011), suggesting that infection time longer than the average and reactivation events could be associated with the accumulation of genetic diversity. The characteristics of the healthcare system, the efficiency of the drug treatment and the socioeconomic status also have a role on the genetic diversity expected. The analysis of diversity in the population of MTB has been found to be associated in certain cases with substandard life conditions, delayed diagnosis and improper therapeutical management and poor adherence to treatment, factors which can be responsible for prolonged infections (Shamputa et al., 2006; Navarro et al., 2011).

Post-mortem genetic analysis of the MTB population within individuals (Cohen et al., 2011) has demonstrated the accumulation of diversity by MTB after a prolonged infection period. It has been shown with special exhaustivity in a recent article by Lieberman et al. (2016), based on post-mortem multiple

sampling and WGS analysis in patients who had received only minimal anti-TB treatment.

The identification of genetic diversity is not a meaningless finding and different authors have found an effect of certain subtle genetic modifications and the expression of neighboring genes and even in the infectivity between closely related variants and the potential functional role of microevolution has been explored (Soto et al., 2004; Tantivitayakul et al., 2010; Pérez-Lago et al., 2011, 2013).

Therefore, although the similarity thresholds defined in the genomic epidemiology of TB (Walker et al., 2013) are based on a wide and solid analysis of patients, additional analysis of microevolution in MTB in specific clinical/epidemiological situations that might represent greater opportunities to acquire variability, which were not extensively represented in the studies defining these cutoffs, could help us to evaluate the robustness of the similarity thresholds proposed and to understand the influence of various infection scenarios on microevolution dynamics.

## MATERIALS AND METHODS

### Genotyping by MIRU-VNTR Analysis

24-Locus MIRU-VNTR multiplex analysis was performed from cultured isolates. DNA was purified using the Qiagen DNA MiniKit (Qiagen, Hilden, Germany).

The final reaction mixture (50 µl) included 25 µl of PCR Master Mix (QIAGEN multiplex PCR kit), 5 µl of Q solution, and 0.25 µM of each labeled and unlabeled oligonucleotide (3.9 µM for loci 4156 and 2059). The primers used for PCR amplification and PCR conditions have been reported elsewhere (Supply et al., 2006; Oelemann et al., 2007). PCR products were sized using capillary electrophoresis in an ABI Prism 3100 genetic analyzer (Applied Biosystems, NLLab Centraal B.V., Haarlem, The Netherlands). MIRU-VNTR types were compared using Bionumerics (4.6 Applied Maths, Sint-Martens-Latem, Belgium).

### Whole-Genome Analysis

DNA was extracted, after recovering all the bacteria present in the liquid cultures by centrifugation, using the standard cetyl trimethyl ammonium bromide (CTAB) method, and DNA libraries were generated following the Nextera XT Illumina protocol (Nextera XT Library Prep kit [FC-131-1024]). Library quality and size distribution were checked by running 2 µl on a 2200 TapeStation Bioanalyzer (Agilent Technologies, USA). The libraries were then normalized based on the average fragment size observed, the library concentration was measured using Qubit 2.0 Fluorometer (Life Technologies, US), and the libraries were pooled. Paired-end sequences were obtained using a MiSeq

platform, with an average per base coverage of 87x (range 62x–113).

SNP calling was performed as indicated elsewhere (Pérez-Lago et al., 2014). In summary, after mapping to a hypothetical MTB ancestral genome (identical to H37Rv according to structure but including the maximum-likelihood-inferred ancestral nucleotide positions from a virtual ancestor; Comas et al., 2010), we extracted all variable positions in the strain of interest. In order to avoid false-positive calls, a series of quality filters were applied to data associated with the SNP. First of all, a minimum coverage >20x and mapping quality 20 were required. From all the variants detected, we divide them into homozygous (present at least in 90% of the reads) and heterozygous calls (present in less than 90% of the reads). Only those SNPs in heterozygosity that appear in homozygosity in other member of the cluster, were selected to be included in the analyses. False positive variants could appear due to mapping errors in genome repetitive regions or near indels. So, we filter out these potential errors by omitting from our analysis variants detected in repetitive regions, phages and PE/PPE regions. Also, those variants found near indel areas and in regions with an anomalous accumulation of SNPs (3 or more SNPs in a 10 bp window) were omitted.

Alignments and SNP variants were visualized and checked using the IGV program. Multiple comparisons between the SNPs from different isolates were performed using an in-house script written in R.

Fastq files with the raw data for isolates are deposited (<http://www.ebi.ac.uk>) under accession numbers ERS2016357-ERS2016427 and ERP002297.

The dN/dS ratio was calculated using the total number of synonymous and non-synonymous variants for the intra- and inter-patient groups. The potential synonymous and non-synonymous sites for the *M. tuberculosis* genome were obtained using the SNAP tool (Ota and Nei, 1994). All the regions omitted in the variant calling pipeline (repeats, phages and PE/PPE) were also omitted when calculating the dN/dS ratio. The reference dataset used to compare the dN/dS ratio was obtained from a previously published work (12808 synonymous SNPs and 21118 non-synonymous SNPs; Comas et al., 2013). Fisher's exact test was applied in the comparison of nonsynonymous/synonymous ratio.

The median-joining networks were constructed from the SNP matrix generated for each case using NETWORK 5.0.0.1.

## RESULTS

We aimed to select representative examples to illustrate deviations in the standard infective circumstances faced by MTB that could provide MTB with better opportunities to acquire variability. We examined both inpatient scenarios (single infections in selected cases) and interpatient scenarios (transmission chains involving serial hosts). We selected 2 settings in which we could expect a more marked inpatient microevolution: prolonged infections due to poor adherence to treatment and cases in which reactivation occurred sometime

after the microbiological cure of the first episode. For the interpatient evaluation, extensive active transmission clusters involving a high number of cases and clusters involving both latency and active transmission events were selected to attempt to obtain a selective pressure that was higher than average.

### Inpatient Analysis: Prolonged Infections

We selected 4 cases involving poor adherence to treatment in which 2 sequential MTB isolates (**Supplementary Table 1**, cases 1–4; all pansusceptible) had been obtained for the same episode over a longer period than usual (20–29 months). Reinfection was excluded because the paired isolates shared identical MIRU-VNTR patterns (data not shown). The comparative WGS analysis of the paired isolates revealed the occurrence of variability, with a range of 1–3 SNPs between isolates (**Figure 1A**; **Supplementary Table 1**).

### Inpatient Analysis: Reactivations

Two cases with complete adherence to treatment and a second isolate with an identical MIRU-VNTR pattern obtained 27 and 56 months after the resolution of the first episode were selected as representatives of reactivation. We evaluated whether the stress arising from the transition from latency to reactivation could lead to greater accumulation of variability. Microevolution was identified in both cases, and 1 and 5 SNPs were found between the sequential isolates (**Supplementary Table 1**, cases 5 and 6) (**Figure 1B**).

### Interpatient Analysis: Extensive Active Transmission Clusters

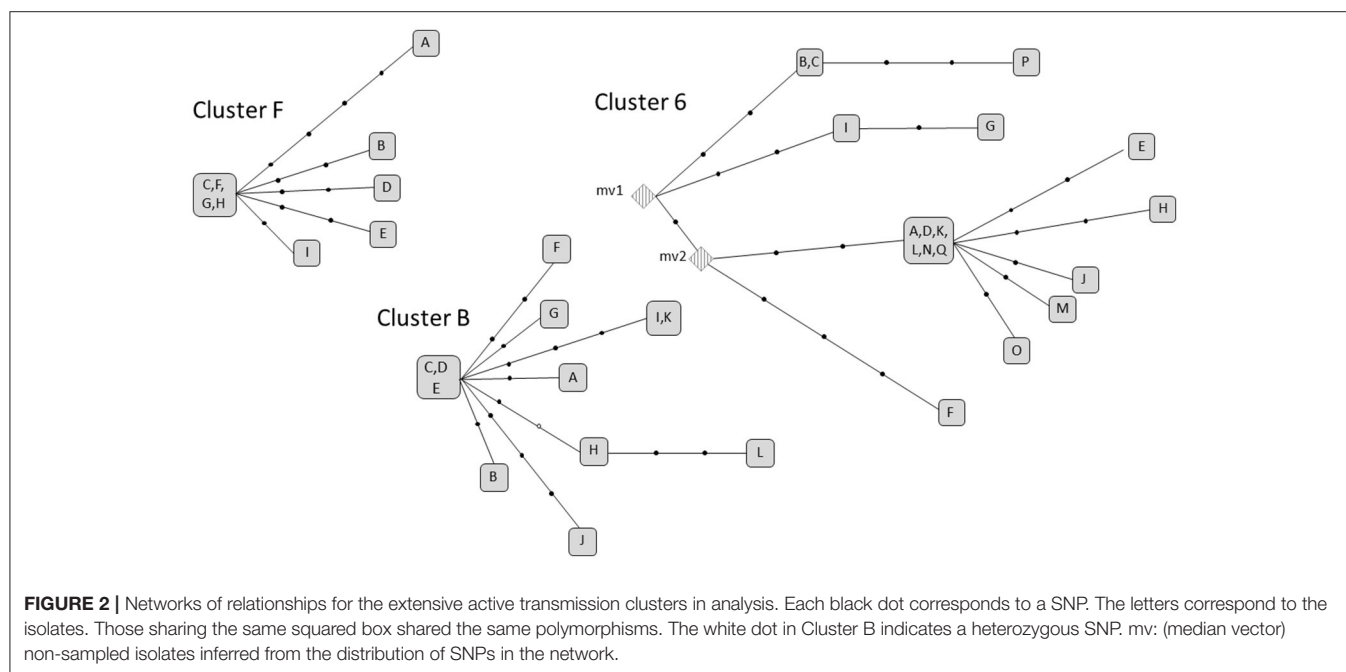
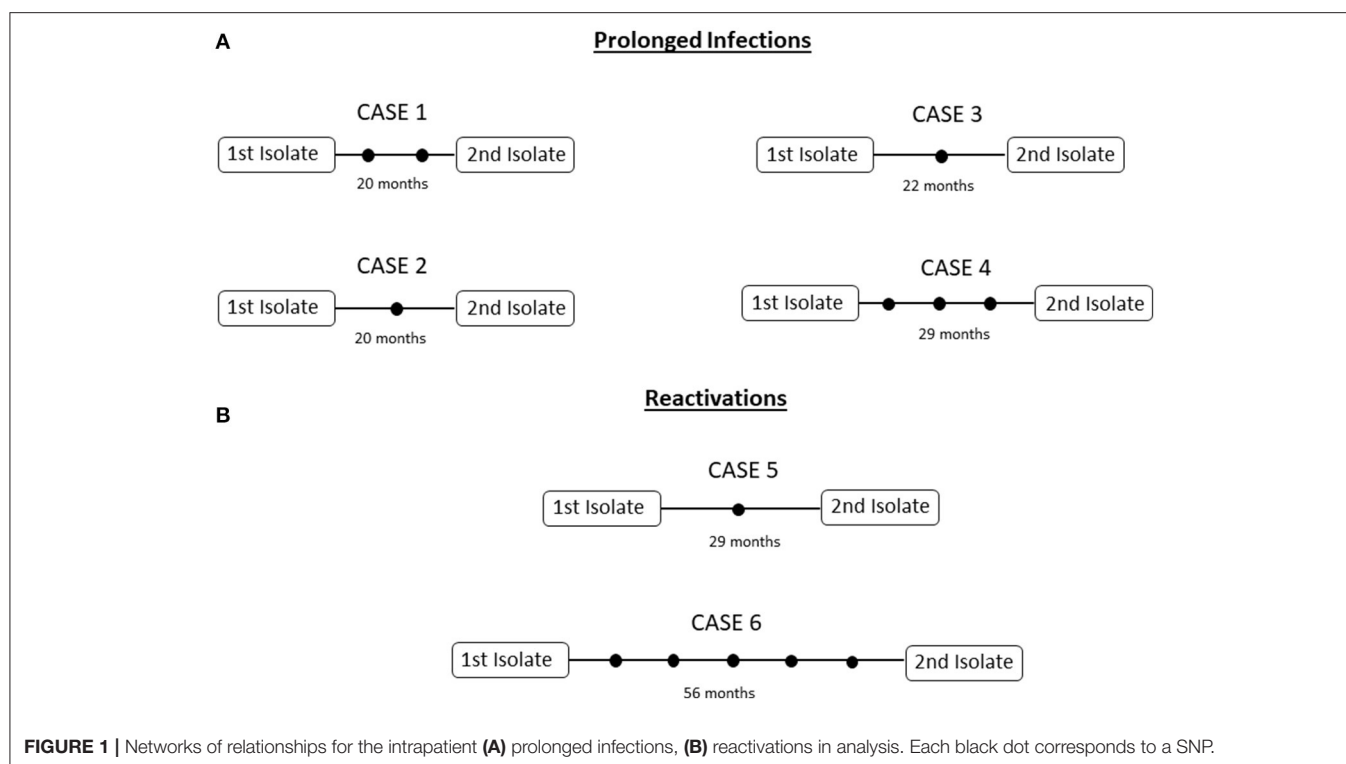
We selected 3 clusters (cluster F, 6 and B) that fulfilled the double criteria of prolonged infection (6, 11, and 12 years, respectively) and a high number of cases (9–17 cases). In all cases, the strains involved were pansusceptible. The microevolution observed in each cluster led to accumulation of, i.e., 11, 20, and 15 SNPs (one of which was in heterozygosity), respectively (**Supplementary Tables 2–4**). However, the accumulation dynamics of SNPs between each 2 linked cases in the transmission networks was moderate (**Figure 2**), with the maximum distances between any of 2 clustered isolates in clusters 6, B, and F being 9, 7, and 6.

### Interpatient Analysis: Clusters Including Both Latency and Active Transmission

We now aimed to analyze events with coincidental involvement of latency and active transmission through sequential hosts and added the requirements of prolonged observation periods and/or the involvement of a high number of cases. We identified 3 clusters in 3 different geographic settings (Cluster 1 in Madrid, Spain; Cluster 2 in Sabadell, Spain; and Cluster 3 in Latgale, Latvia) fulfilling as many coincidental factors as possible (see above).

Cluster 1 corresponded to an 11-case outbreak in a school in Madrid. The event consisted of a double outbreak involving the same setting but 3 years apart (**Figure 3**). In 2012, 2 MTB isolates with identical MIRU-VNTR patterns were obtained from

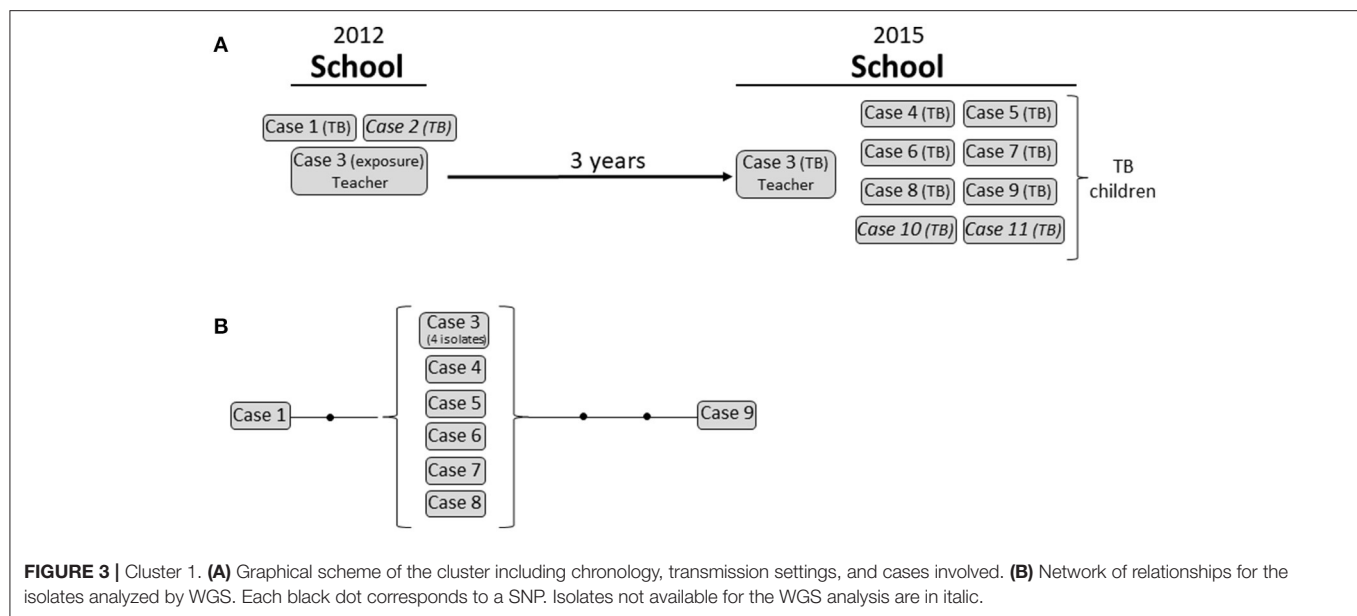




2 epidemiologically linked cases (a child from the school and the father of a child from another class). At the time, contact tracing was incomplete. One of the teachers (case 3)—likely exposed in the 2012 event—developed active TB 3 years later (2015), and 8 new cases of culture-positive TB were diagnosed among the children in the school. All 11 isolates were identical according

to the MIRU-VNTR analysis. Unfortunately, 3 isolates were not available for WGS, although we included 4 isolates from the teacher to study inpatient variability.

We expected that the combination of these coincidental factors (latency, active transmission, and multi-host infection) in this complex event would reveal marked variability between



the strains involved. However, global variability in SNPs was limited (**Figure 3; Supplementary Table 5**). All 4 isolates from the teacher and 5 out of 6 isolates from 2015 were identical (0 SNPs) and showed only 1 SNP of difference compared with the 2012 isolate (case 1). One of the isolates from 2015 (case 9) showed 2 additional SNPs.

Cluster 2 corresponded to a complex family outbreak in Sabadell involving 9 cases and spanning 6 years owing to the overlap of active transmission events and latency periods between the exposure of specific cases where the disease developed some years later and eventually coincided with 3 reactivations (**Figure 4**). Seven isolates were available for study, and all of them were identical according to MIRU-VNTR analysis. Despite the complexity of the cluster, WGS revealed global variability for only 5 SNPs (**Figure 4; Supplementary Table 5**).

Cluster 3 corresponded to a school and household outbreak in Latvia (**Figure 4**) that involved 5 cases with active TB [1 child (index case), her 3 sisters, and 1 adult (the teacher)] in 2012. Three cases exposed in 2012 developed the disease 1 year later (case 6) and 2 years later (cases 8 and 9). The scenario was further complicated when one of the 2012 cases (case 4) reactivated 2 years later and transmitted TB to her husband (case 7). All 10 isolates were available and shared an identical MIRU-VNTR pattern. WGS analysis revealed that all but 1 of the cases was identical (0 SNPs); however, in the remaining case (case 2), we observed a burst of variability that led to the accumulation of 5 SNPs (**Figure 5; Supplementary Table 5**).

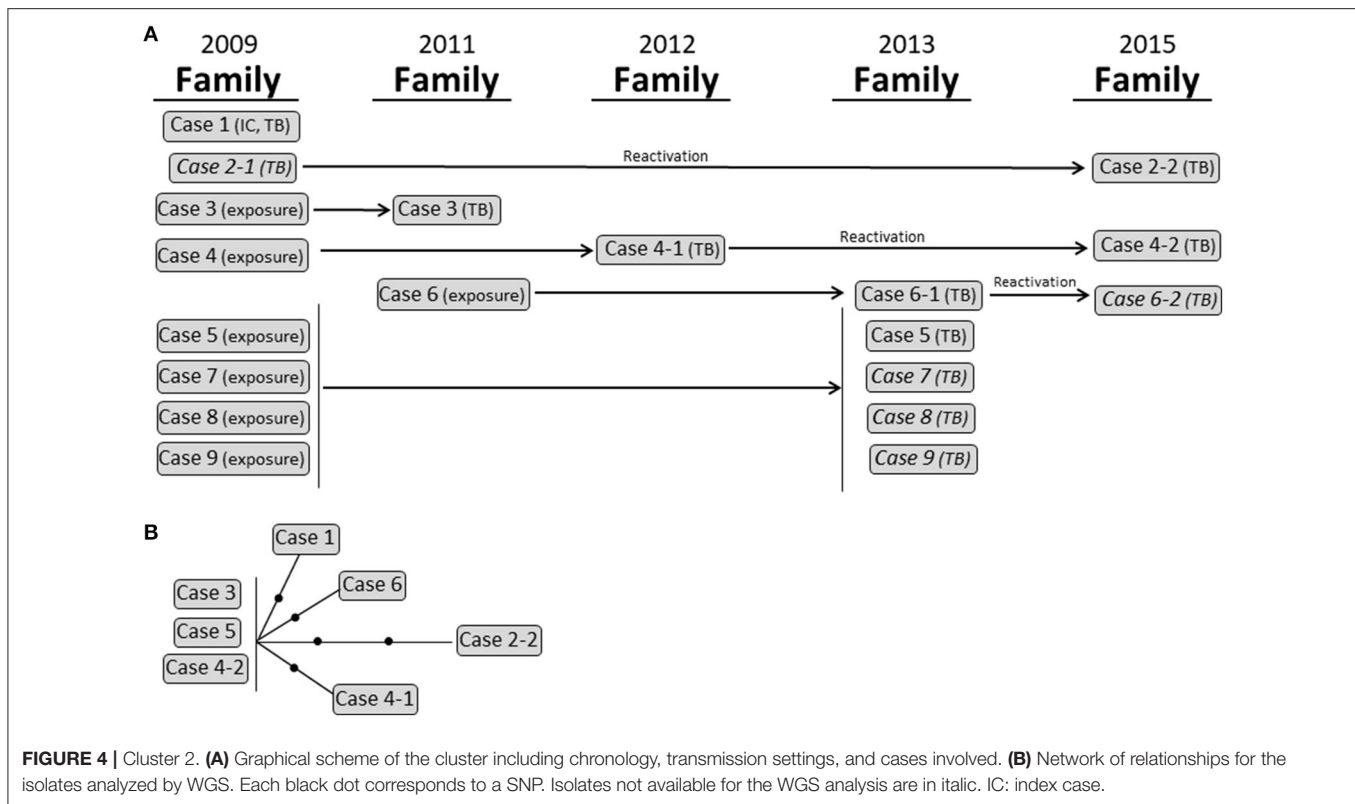
## DISCUSSION

Microevolution is a phenomenon which can be modulated by various factors. We have to consider it at least a bifactorial event in which both bacterial and host factors intervene. The role of bacterial factors is illustrated by the finding that certain strains from the Beijing lineage show a hypermutator phenotype

(Ebrahimi-Rad et al., 2003). Other non-Beijing strains have been also described to be prone to acquire diversity (Navarro et al., 2017). Microevolution events leading to higher than expected accumulation of variability have been partially addressed in a previous study from our group (Pérez-Lago et al., 2014). Some of these events involved sequential host-to-host transmissions, but an equivalent degree of diversity was observed also within individuals. This inpatient diversity was detected between isolates infecting different organs, in patients with respiratory and extrapulmonary infection, but also in isolates recovered strictly from the respiratory site. These data mean that the inpatient diversity might impact on the inference of recent transmission clusters (Pérez-Lago et al., 2014).

As our previous study was based on a convenience sample of isolates selected once they had been shown to have microevolved, we now aimed to perform a more systematic evaluation (i.e., with no preselection of microevolved isolates) of the impact of several circumstances on the acquisition of variability that MTB infection could meet in various clinical or epidemiological scenarios. As with any evolutionary process, microevolution in MTB is the combination of the opportunities the microorganism has to acquire variability and the occurrence of selective pressure bottlenecks that force the selection of clonal variants from among those which emerge. Following this rationale, we selected clinical/epidemiological situations that could lead to increased acquisition of variability (both inpatient and interpatient).

We began with the inpatient scenario by selecting single case infections, looking for situations with longer periods of active infection (prolonged infections) in order to increase the likelihood of the emergence of variants and/or looking for factors involved in selective pressure (such as intermittent poor adherence to treatment or transitions from latency to active infection in reactivations). In general, our findings highlight the limited variability acquired, which was within



the ranges expected. However, in 1 reactivation (case 6, **Supplementary Table 1**), we observed greater accumulation of variability (5 SNPs), thus reaching the lower threshold for considering 2 cases as not epidemiologically linked (Walker et al., 2013). Previously, we had also observed inpatient situations leading to a higher accumulation of SNPs (Pérez-Lago et al., 2014; Navarro et al., 2017).

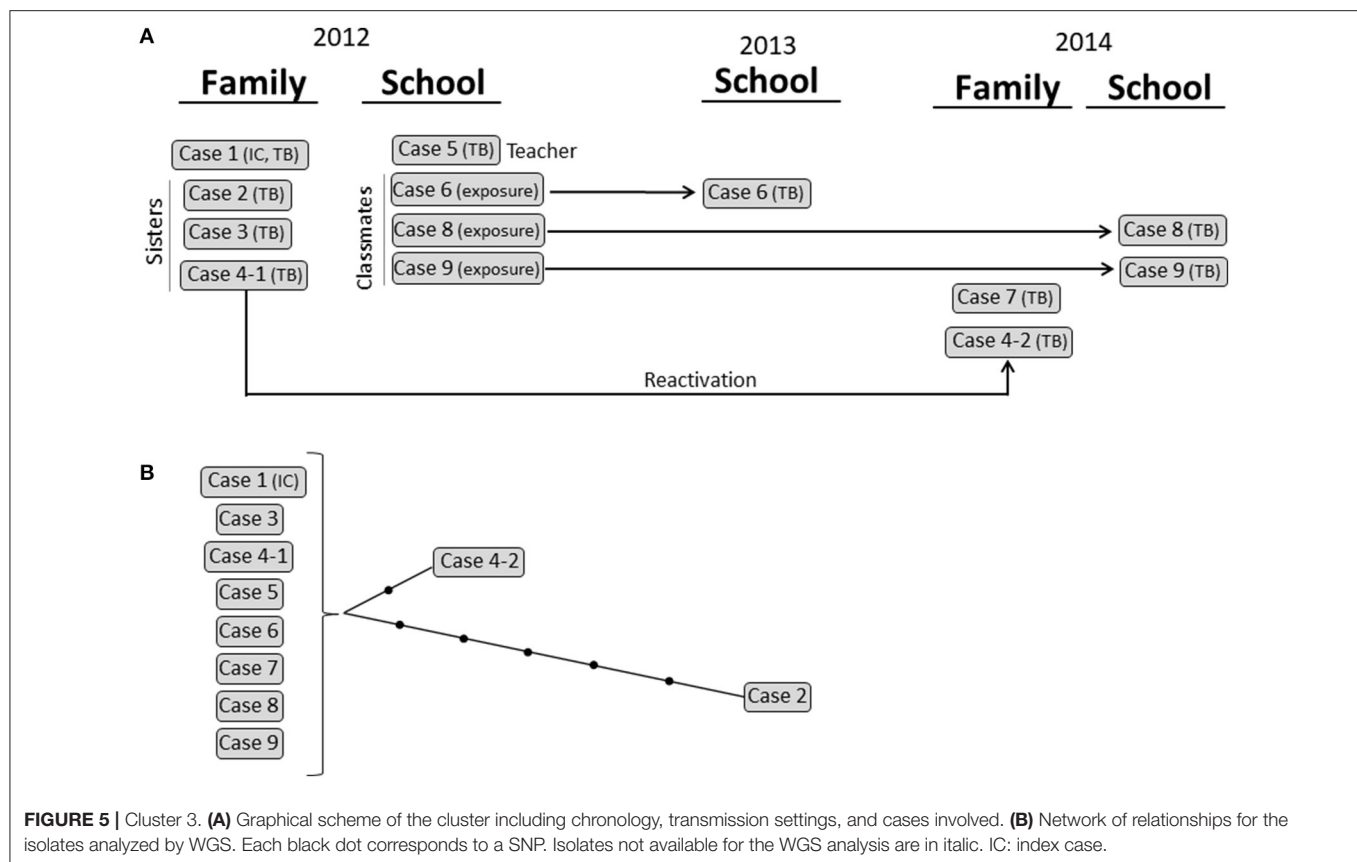
The non-systematic association between a greater accumulation of variability and clinical situations that could theoretically favor it underlines the multifactorial nature of this phenomenon. Observations from other studies also lead to contradictory findings. Whereas no variability is found in circumstances where it is highly likely, such as long-term persistent infection by a Beijing strain with constant intermittent treatment (Pérez-Lago et al., 2015a), marked variability is detected in less extreme situations (Pérez-Lago et al., 2014; Navarro et al., 2017). It is also true that greater diversity is described in studies involving MDR strains in which a more dynamic process of acquisition of variability is expected owing to constant competition between resistant and compensatory mutations (Merker et al., 2013), although competition between clonal variants is also expected in heterogeneous populations involving susceptible strains.

We then moved on to the interpatient analysis, which focused first on extensive clusters to evaluate the acquisition of variability when 2 key factors coincided, namely, prolonged observation time and adaptation to a high number of sequentially infected hosts. However, the maximum genetic distance accumulated never surpassed 9 SNPs, and the distance between sequential

isolates ranged from 1 to 6 SNPs, i.e., within the previously established thresholds although approaching in some cases to their limits (Walker et al., 2013).

We then looked for the coincidence of several of the factors that had been evaluated independently, making it possible to observe amplification of effects that could only have a subtle impact when evaluated one by one. It was not easy to find transmission events with the intervention of as many as possible of these factors, namely, extended time to allow microevolution to occur, involvement of a high number of hosts to ensure sequential selective pressures due to the establishment of each infection in different hosts, and overlap of transitions from latency to active infections. These kind of clusters with such a complex epidemiological peculiarities were insufficiently represented in the studies defining the SNP thresholds to be applied as a reference in genomic epidemiology (Walker et al., 2013). We found 3 clusters fulfilling all or most of our requirements, each from a different population and involving 2 countries. However, despite the greater opportunities to acquire diversity that the clinical/epidemiological situations in these complex clusters offered, the global variability identified generally lay within the thresholds established, which limits the impact of this diversity on the inference of recent transmission in genomic epidemiology studies.

Although our main aim was to perform a quantitative analysis of the number of SNPs acquired, we had the opportunity to record interesting qualitative observations. The first was the finding that some SNPs (1 each in case 4 and case 6, **Supplementary Table 1**; and one each in case 1 and 4-1 in



cluster 2, **Supplementary Table 5**) were only observed in the first isolate of a chronological series. While this observation may at first sight seem incongruent, owing to the negligible degree of homoplasmy described for *Mycobacterium tuberculosis complex* (Hershberg et al., 2008; Comas et al., 2009) and the highly unlikely reversion of mutations once acquired, we believe that it might be explained by undetected coexistence of the variant which acquired the SNP together with the parental variant, which did not have the SNP. Given the unavoidable limitations of sputum sampling with respect to the whole bacterial population in the lung, we likely only detected 1 of the coexisting variants at each sampling point. The observation that isolates recovered from 1 sputum sample do not represent true clonal complexity in the lung has been addressed elsewhere (Black et al., 2015; Pérez-Lago et al., 2015b).

About the functional significance of SNPs (13 synonymous and 40 non-synonymous SNPs), it must be highlighted a lower nonsynonymous/synonymous ratio (0.325) compared with a global database including 220 strains (0.606,  $p = 0.04$  Fisher's exact test). The difference is mainly driven by the accumulation of a lower number of nonsynonymous changes in inpatient cases compared to inter-patient transmission cluster (dN/dS 0.353 vs. 1.714). This observation is in contrast with the expectation that shorter timespan should be associated to higher dN/dS ratios. The lower number of nonsynonymous changes inpatient may indicate the action of purifying selection. Purifying selection

is expected during treatment course (Trauner et al., 2017) and can remove diversity faster, given the clonal nature of MTBC populations. However, the heterogeneity of the dataset involving different patients and transmission clusters as well as the low number of mutations prevents us to support any general conclusion.

Our global analysis did not enable us to identify a clinical/epidemiological situation that could predict greater accumulation of diversity in MTB. We must admit certain limitations resulting of not having considered the diversity based on structural variants and indels or not having performed a high-depth analysis or minority variants. We have focused on diversity based on SNPs mainly due to the fact that this is the criterion in which genomic epidemiology lies, and thus to evaluate the impact of this variability on the inference of recent transmission. The number of SNPs identified under circumstances that could theoretically favor microevolution was limited, within the ranges expected according to the proposed similarity thresholds for considering 2 isolates as related or unrelated (Walker et al., 2013). These findings support the robustness of the cutoffs. However, we also found asymmetric bursts of variability, as reported elsewhere (Pérez-Lago et al., 2014; Navarro et al., 2017). The coincidence of several such cases in the same transmission event could eventually lead us to overlook some transmission clusters.

We must also accept that despite trying to force the coincidence of factors theoretically favoring microevolution,



some of these factors, such as length of the active infection period, serial involvement of sequentially infected hosts due to active recent transmission, and lack of adherence to treatment, could be more frequent and of a greater magnitude in settings with weaker diagnostic and control programs. Studies similar to ours should be replicated in these settings to enrich our knowledge of the circumstances which could favor diversity in TB and to determine more precisely its impact on the tracking of transmission.

## AUTHOR CONTRIBUTIONS

Experimental load: all authors; Writing: LP-L, DG-d-V; Design conceptualization: LP-L, DG-d-V; Bioinformatic Analysis and design: ÁC-O, IC.

## ACKNOWLEDGMENTS

We thank Thomas O'Boyle for proofreading the manuscript. This study was funded by a grant from the Ministry of Economy and Competitiveness ISCIII FIS (grant 15/01554) and cofunded by ERDF (FEDER) Funds from the European Commission: "A way of making Europe." IP and IO were

supported by the Latvian National Research program VPP 5.7 "BIOMEDICINE." LP-L was supported by M. Servet contract MS15/00075 and CP15/00075. IC was supported by Ministerio de Educación y Ciencia (Ref SAF2016-77346-R) and the European Research Council (ERC) (638553-TB-ACCELERATE). AC-O is recipient of a FPU fellowship (FPU13/00913) from Ministerio de Educación y Ciencia (Spanish Government).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02661/full#supplementary-material>

**Supplementary Table 1** | SNPs and features for the SNPs found for the inpatient analysis.

**Supplementary Table 2** | SNPs and features for the SNPs found for the Cluster F.

**Supplementary Table 3** | SNPs and features for the SNPs found for the Cluster 6.

**Supplementary Table 4** | SNPs and features for the SNPs found for the Cluster B.

**Supplementary Table 5** | SNPs and features for the SNPs found for the Clusters 1, 2 and 3.

## REFERENCES

- Al-Hajjaj, S. A., Akkerman, O., Parwati, I., al-Gamdi, S., Rahim, Z., van Soolingen, D., et al. (2010). Microevolution of *Mycobacterium tuberculosis* in a tuberculosis patient. *J. Clin. Microbiol.* 48, 3813–3816. doi: 10.1128/JCM.00556-10
- Black, P. A., de Vos, M., Louw, G. E., van der Merwe, R. G., Dippenaar, A., Streicher, E. M., et al. (2015). Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC Genomics* 16, 857. doi: 10.1186/s12864-015-2067-2
- Casali, N., Broda, A., Harris, S. R., Parkhill, J., Brown, T., and Drobniewski, F. (2016). Whole genome sequence analysis of a large isoniazid-resistant tuberculosis outbreak in London: a retrospective observational study. *PLoS Med.* 13:e1002137. doi: 10.1371/journal.pmed.1002137
- Cohen, T., Wilson, D., Wallengren, K., Samuel, E. Y., and Murray, M. (2011). Mixed-strain *Mycobacterium tuberculosis* infections among patients dying in a hospital in KwaZulu-Natal, South Africa. *J. Clin. Microbiol.* 49, 385–388. doi: 10.1128/JCM.01378-10
- Comas, I. (2017). Genomic epidemiology of tuberculosis. *Adv. Exp. Med. Biol.* 1019, 79–93. doi: 10.1007/978-3-319-64371-7\_4
- Comas, I., Chakravarti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., et al. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* 42, 498–503. doi: 10.1038/ng.590
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., et al. (2013). Out-of-Africa migration and neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45, 1176–1182. doi: 10.1038/ng.2744
- Comas, I., Homolka, S., Niemann, S., and Gagneux, S. (2009). Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE* 4:e7815. doi: 10.1371/journal.pone.0007815
- Ebrahimi-Rad, M., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., et al. (2003). Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerging Infect. Dis.* 9, 838–845. doi: 10.3201/eid0907.020803
- Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., et al. (2008). High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* 6:e311. doi: 10.1371/journal.pbio.0060311
- Lieberman, T. D., Wilson, D., Misra, R., Xiong, L. L., Moodley, P., Cohen, T., et al. (2016). Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nat. Med.* 22, 1470–1474. doi: 10.1038/nm.4205
- Merker, M., Kohl, T. A., Roetzer, A., Truebe, L., Richter, E., Rüscher-Gerdes, S., et al. (2013). Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PLoS ONE* 8:e82551. doi: 10.1371/journal.pone.0082551
- Navarro, Y., Herranz, M., Pérez-Lago, L., Martínez Lirola, M., Indal, T. B., Ruiz-Serrano, M. J., et al. (2011). Systematic survey of clonal complexity in tuberculosis at a populational level and detailed characterization of the isolates involved. *J. Clin. Microbiol.* 49, 4131–4137. doi: 10.1128/JCM.05203-11
- Navarro, Y., Pérez-Lago, L., Herranz, M., Sierra, O., Comas, I., Sicilia, J., et al. (2017). In-depth characterization and functional analysis of clonal variants in a *Mycobacterium tuberculosis* strain prone to microevolution. *Front. Microbiol.* 8:694. doi: 10.3389/fmicb.2017.00694
- Nikolayevskiy, V., Kranzer, K., Niemann, S., and Drobniewski, F. (2016). Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: a systematic review. *Tuberculosis* 98, 77–85. doi: 10.1016/j.tube.2016.02.009
- Oelemann, M. C., Diel, R., Vatin, V., Haas, W., Rüscher-Gerdes, S., Locht, C., et al. (2007). Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J. Clin. Microbiol.* 45, 691–697. doi: 10.1128/JCM.01393-06
- Ota, T., and Nei, M. (1994). Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. *Mol. Biol. Evol.* 11, 613–619.
- Pérez-Lago, L., Comas, I., Navarro, Y., Gonzalez-Candelas, F., Herranz, M., Bouza, E., et al. (2014). Whole genome sequencing analysis of inpatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J. Infect. Dis.* 209, 98–108. doi: 10.1093/infdis/jit439
- Pérez-Lago, L., Herranz, M., Lirola, M. M., Group, I.-T., Bouza, E., and García de Viedma, D. (2011). Characterization of microevolution events in *Mycobacterium tuberculosis* strains involved in recent transmission clusters. *J. Clin. Microbiol.* 49, 3771–3776. doi: 10.1128/JCM.01285-11

- Pérez-Lago, L., Navarro, Y., Herranz, M., Bouza, E., and García-de-Viedma, D. (2013). Differences in gene expression between clonal variants of *Mycobacterium tuberculosis* emerging as a result of microevolution. *Int. J. Med. Microbiol.* 303, 674–677. doi: 10.1016/j.ijmm.2013.09.010
- Pérez-Lago, L., Navarro, Y., Montilla, P., Comas, I., Herranz, M., Rodríguez-Gallego, C., et al. (2015a). Persistent infection by a *Mycobacterium tuberculosis* strain that was theorized to have advantageous properties, as it was responsible for a massive outbreak. *J. Clin. Microbiol.* 53, 3423–3429. doi: 10.1128/JCM.01405-15
- Pérez-Lago, L., Palacios, J. J., Herranz, M., Ruiz Serrano, M. J., Bouza, E., and García-de-Viedma, D. (2015b). Revealing hidden clonal complexity in *Mycobacterium tuberculosis* infection by qualitative and quantitative improvement of sampling. *Clin. Microbiol. Infect.* 21, 147.e1–7. doi: 10.1016/j.cmi.2014.09.015
- Satta, G., Lipman, M., Smith, G. P., Arnold, C., Kon, O. M., and McHugh, T. D. (2017). *Mycobacterium tuberculosis* and Whole-genome sequencing: how close are we to unleash its full potential? *Clin. Microbiol. Infect.* doi: 10.1016/j.cmi.2017.10.030. [Epub ahead of print].
- Seto, J., Wada, T., Suzuki, Y., Ikeda, T., Mizuta, K., Yamamoto, T., et al. (2017). *Mycobacterium tuberculosis* transmission among elderly persons, Yamagata prefecture, Japan, 2009–2015. *Emerging Infect. Dis.* 23, 448–455. doi: 10.3201/eid2303.161571
- Shamputa, I. C., Jugheli, L., Sadradze, N., Willery, E., Portaels, F., Supply, P., et al. (2006). Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. *Respir. Res.* 7:99. doi: 10.1186/1465-9921-7-99
- Soto, C. Y., Menéndez, M. C., Pérez, E., Samper, S., Gómez, A. B., García, M. J., et al. (2004). IS6110 mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. *J. Clin. Microbiol.* 42, 212–219. doi: 10.1128/JCM.42.1.212-219.2004
- Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., et al. (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 44, 4498–4510. doi: 10.1128/JCM.01392-06
- Tantivitayakul, P., Panapruksachai, S., Billamas, P., and Palittapongarnpim, P. (2010). Variable number of tandem repeat sequences act as regulatory elements in *Mycobacterium tuberculosis*. *Tuberculosis* 90, 311–318. doi: 10.1016/j.tube.2010.08.003
- Trauner, A., Liu, Q., Via, L. E., Liu, X., Ruan, X., Liang, L., et al. (2017). The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome Biol.* 18:71. doi: 10.1186/s13059-017-1196-0
- Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., et al. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* 13, 137–146. doi: 10.1016/S1473-3099(12)70277-3
- Warren, R. M., van der Spuy, G. D., Richardson, M., Beyers, N., Booysen, C., Behr, M. A., et al. (2002). Evolution of the IS6110-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 40, 1277–1282. doi: 10.1128/JCM.40.4.1277-1282.2002
- Witney, A. A., Bateson, A. L., Jindani, A., Phillips, P. P., Coleman, D., Stoker, N. G., et al. (2017). Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. *BMC Med.* 15:71. doi: 10.1186/s12916-017-0834-4

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Herranz, Pole, Ozere, Chiner-Oms, Martínez-Lirola, Pérez-García, Gijón, Serrano, Romero, Cuevas, Comas, Bouza, Pérez-Lago and García-de-Viedma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# In-Depth Characterization and Functional Analysis of Clonal Variants in a *Mycobacterium tuberculosis* Strain Prone to Microevolution

Yurena Navarro<sup>1,2,3,4,5</sup>, Laura Pérez-Lago<sup>1,2,3</sup>, Marta Herranz<sup>1,2,3</sup>, Olalla Sierra<sup>1,2</sup>, Iñaki Comas<sup>6,7</sup>, Javier Sicilia<sup>2,8</sup>, Emilio Bouza<sup>1,2,3,9</sup> and Darío García de Viedma<sup>1,2,3,4\*</sup>

<sup>1</sup> Servicio Microbiología Clínica y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón, Madrid, Spain, <sup>2</sup> Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain, <sup>3</sup> CIBER Enfermedades Respiratorias, CIBERES, Madrid, Spain, <sup>4</sup> CEI Campus Moncloa, UCM-UPM, Madrid, Spain, <sup>5</sup> Centro de Vigilancia Sanitaria Veterinaria, Universidad Complutense Madrid, Madrid, Spain, <sup>6</sup> Unidad Mixta Genómica y Salud, Centro Superior de Investigación en Salud Pública (FISABIO)-Universitat de València, Valencia, Spain, <sup>7</sup> CIBER en Epidemiología y Salud Pública, Madrid, Spain, <sup>8</sup> Unidad de Medicina y Cirugía Experimental, Hospital General Universitario Gregorio Marañón, Madrid, Spain, <sup>9</sup> Departamento de Medicina, Facultad de Medicina, Universidad Complutense de Madrid, Madrid, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Anders Norman,  
Statens Serum Institut, Denmark  
Igor Mokrousov,  
Saint Petersburg Pasteur Institute,  
Russia

### \*Correspondence:

Darío García de Viedma  
dgvedma2@gmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 22 February 2017

Accepted: 04 April 2017

Published: 24 April 2017

### Citation:

Navarro Y, Pérez-Lago L, Herranz M,  
Sierra O, Comas I, Sicilia J, Bouza E  
and García de Viedma D (2017)  
In-Depth Characterization and  
Functional Analysis of Clonal Variants  
in a *Mycobacterium tuberculosis*  
Strain Prone to Microevolution.  
Front. Microbiol. 8:694.  
doi: 10.3389/fmicb.2017.00694

The role of clonal complexity has gradually been accepted in infection by *Mycobacterium tuberculosis* (MTB), although analyses of this issue are limited. We performed an in-depth study of a case of recurrent MTB infection by integrating genotyping, whole genome sequencing, analysis of gene expression and infectivity in *in vitro* and *in vivo* models. Four different clonal variants were identified from independent inpatient evolutionary branches. One of the single-nucleotide polymorphisms in the variants mapped in *mce3R*, which encodes a repressor of an operon involved in virulence, and affected expression of the operon. Competitive *in vivo* and *in vitro* co-infection assays revealed higher infective efficiency for one of the clonal variants. A new clonal variant, which had not been observed in the clinical isolates, emerged in the infection assays and showed higher fitness than its parental strain. The analysis of other patients involved in the same transmission cluster revealed new clonal variants acquired through novel evolutionary routes, indicating a high tendency toward microevolution in some strains that is not host-dependent. Our study highlights the need for integration of various approaches to advance our knowledge of the role and significance of microevolution in tuberculosis.

**Keywords:** microevolution, *Mycobacterium tuberculosis*, functional analysis, *in vitro* infections, *in vivo* infections, whole genome sequencing

## INTRODUCTION

The idea that *Mycobacterium tuberculosis* (MTB) infection of a single case is caused by a single strain is increasingly debated. Genotyping has enabled us to describe cases of co-infection by  $\geq 2$  different strains (mixed infection) and coexistence of clonal variants of the same strain (polyclonal infection) (García de Viedma et al., 2005; Shamputa et al., 2006; Al-Hajj et al., 2010; Navarro et al., 2011; Huyen et al., 2012; Muwonge et al., 2013). In the latter, clonal variants emerge through microevolution phenomena, which make it possible to detect subtle changes when applying standard fingerprinting strategies (IS6110-RFLP or mycobacterial interspersed repetitive

units–variable number of tandem repeats [MIRU-VNTR]). These changes can have a functional impact on the expression of neighboring genes and can increase the variability of antigenic proteins (Akhtar et al., 2009; Olsen et al., 2009; Yindeeyoungyeon et al., 2009; Tantivitayakul et al., 2010; Perez-Lago et al., 2011, 2013).

Most studies on microevolution in infection by MTB or on the functional characterization of MTB clonal variants merely describe the genotypic changes involved. If they include a functional analysis, then this is only partially based on gene expression assays or standard infection models. In our study, we attempted to overcome this fragmented view of clonal complexity of MTB infection by performing an in-depth analysis of a representative case of clonally complex MTB infection by integrating the following approaches: (i) complete characterization of the variants by both standard genotyping and whole genome sequencing (WGS); (ii) gene expression analysis; and (iii) evaluation of infectivity by analyzing the behavior of clonal variants not only in standard infections, but also in co-infections in broader experimental conditions than usual, i.e., applying both cellular and animal models. We show how integrating various approaches can advance our knowledge of the role and significance of microevolution in MTB infection.

## MATERIALS AND METHODS

### *Mycobacterium tuberculosis* Clonal Variants

The clonal variants (A and B) were isolated from a patient with recurrent tuberculosis and identified using 15-locus MIRU-VNTR, as described in Martín et al. (2011).

### Expanded Characterization of the Clonal Variants

#### 24-Locus MIRU-VNTR

Clonal variants were further characterized using 24-locus MIRU-VNTR (Oelemann et al., 2007), as described in Navarro et al. (2011), and IS6110-restriction fragment length polymorphism (RFLP), following international standardization guidelines (van Embden et al., 1993).

#### Ligation-Mediated PCR (LM-PCR)

IS6110 sequences were mapped using LM-PCR, as described in Perez-Lago et al. (2011).

#### Allele-Specific-PCR (ASO-PCR)

Allele-specific PCRs were designed to analyze single colonies to assign the alleles present for several SNPs identified in the following Rvs: Rv1201, 1527, 1553, 1963, 2209, 2579.

The reactions were carried out in 50  $\mu$ l including MgCl<sub>2</sub> (1.2, 0.9, 1.5, 1.3, 1.3, and 1.3 mM; for the different SNPs in the order indicated above), 0.4  $\mu$ l dNTPs (10  $\mu$ M); 0.5  $\mu$ l DMSO, 1.5  $\mu$ l s of each primer (Supplementary Table; 10  $\mu$ M), 0.4  $\mu$ l AmpliTaq Gold enzyme. The PCR run was constituted by 30 rounds of the cycle: 10' 95°C, 1' 95°C, 1' Tm (64, 60, 63, 64, 64, 64°C, for the different SNPs in the order indicated above), 10' 72°C and a final tail of 4°C.

For each SNP, two independent PCRs were run. The first one using the selective primer complementary for the allele in the SNP and, as a control, a second PCR targeting the complementary allele. The expected sizes for the products were 239, 402, 239, 207, 234, and 147 bp (for the different SNPs in the order indicated above).

### Whole Genome Sequencing

WGS was performed as detailed elsewhere (Perez-Lago et al., 2014) using a HiSeq 2000 device and a Miseq device (Illumina), which generated 101–51–bp paired-end reads. We mapped the reads for each strain using the Burrows-Wheeler Aligner and the ancestral MTB genome as detailed elsewhere (Comas et al., 2013). SNP calls were made with SAMtools and VarScan (coverage of at least 10 $\times$ , mean SNP mapping quality of 20). The genome was compared between strains using an in-house script written in R.

### Gene Expression Assay

Relative quantification assays based on reverse transcriptase polymerase chain reaction (RT-PCR) were performed to examine expression of the gene involved in the microevolution events. The clonal variants were grown in 7H9 liquid medium (Difco) supplemented with 10% ADC (Becton Dickinson) and 1% Tween 80 (Merck) for 3 weeks. Cell lysis and RNA extraction were performed as previously described (Perez-Lago et al., 2013). RNA was reversed-transcribed using the High Capacity RNA-to-cDNA Kit (Invitrogen, Life Technologies, CA, USA). This step was followed by qRT-PCR amplification (preincubation at 95°C for 10 min and 45 cycles of 95°C for 10 s, 60°C for 10 s, and 72°C for 20 s) using the LightCycler® FastStart DNA Master SYBR Green I kit (Roche) and the primers FyrbE3A (5'-GGTGTTCATGCACGTCT-3') and RyrbE3A (5'-CCG ACCGACATGCCCTTATA-3'). The results of the assay were expressed as the ratio of the values obtained for the variant harboring the SNP to the values from the wild-type variant. Three independent quantitative RT-PCR assays were performed using 2 independent RNA extractions. A 1-sample *t*-test was used to determine whether the average expression ratio was statistically different from 1 ( $p < 0.05$ ).

### Infection Assays

#### In vivo Infection Model

Specific pathogen-free 8-week-old female Balb/c mice were obtained from Charles River Laboratories (L'Arbresle, France) and from our animal experimental laboratory (Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain). The mice were shipped under appropriate conditions, with the corresponding certificate of health and origin. All the animals were kept under controlled conditions in a biosafety level 3 facility with food and water *ad libitum*.

Mice were anesthetized by intraperitoneal injection with xylazine (0.75 mg/g) and ketamine (0.1 mg/g) and subsequently infected with 200  $\mu$ l of inoculum (1–5  $\times$  10<sup>5</sup> bacilli) by intravenous inoculation in the lateral tail vein.

Monitoring of infections by each clonal variant was based on colony forming unit (CFU) counts. The lungs and spleen of the infected mice were individually homogenized in PBS



supplemented with 0.05% Tween 80. Serial 10-fold dilutions of the homogenate were plated on Middlebrook 7H11 agar (Difco). The growth rates for each clonal variant were calculated by linear regression of the CFU logarithm. Eighteen mice were infected, and 3 mice were analyzed at each point for each strain (day 1, week 1, and week 3).

The results of the assays in which the mice were simultaneously co-infected with the 2 clonal variants were analyzed by comparing the proportion of co-infecting strains before infection (adjusting to a 1:1 proportion) and after infection. The lungs and spleen of the infected mice were individually homogenized and plated at each time point (day 1 and week 5) to analyze the allelic value of locus MIRU 42 in 40 single colonies. Ten mice were co-infected, and standard deviations were calculated on the basis of the results obtained from 5 mice at each time point.

In vitro Infection Model

THP-1 cells were differentiated to macrophages and simultaneously co-infected with both clonal variants following the protocol described by Alonso et al. (2010) at a multiplicity of infection of 3 bacteria per cell. The proportion of co-infecting variants before and after infection was calculated at 3 h and day 7 by plating 10-fold serial dilutions of lysates on Middlebrook 7H11 agar (Difco). Seventy colonies were analyzed at each time point using simplex PCR of locus MIRU 42.

Fitness Assay

Clonal variants were subcultured on Mycobacteria Growth Indicator Tubes (BACTEC MGIT 960 System; Becton Dickinson) supplemented with BBL™ PANTA™ and BACTEC MGIT 960 Growth Supplement, as indicated by the manufacturer. Inocula were obtained from positive tubes (3 days) following the protocol used for antimicrobial susceptibility testing. Three MGITs were inoculated after ensuring equivalent bacterial concentrations in the 3 aliquots of inoculum used by plating them on Middlebrook 7H11 agar (Difco). Growth curves were obtained by monitoring the growth units (GU) every hour using BD EpiCenter™. The fitness of the clonal variants was compared based on 2 parameters taken from the growth curves: (i) lag phase (time to a positive

threshold [75 GU]) and (ii) rate growth (time required for the 4,000–6,000–GU increase).

The means and standard deviations were determined, and one-way analysis of variance with repeated measures was used to determine *P*-values, which were adjusted using a Bonferroni correction.

RESULTS

Complete Characterization of Clonal Variants

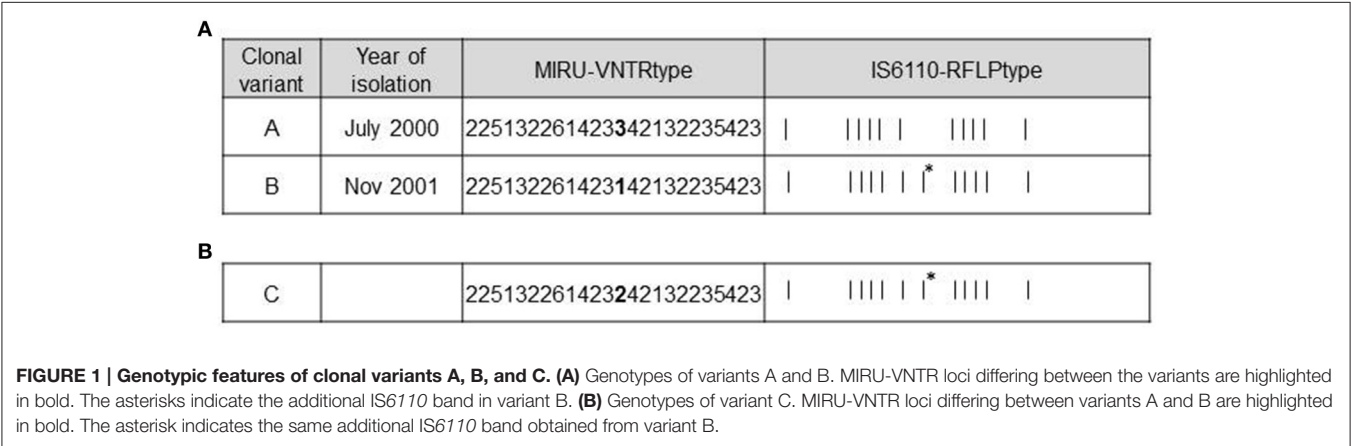
Four isolates from the patient with recurrent infection were available for analysis, 1 for the first episode (July 2000) and 3 for the second one (August–November 2001). Available genotyping data for the clonal variants using 15-locus MIRU-VNTR were further analyzed using 24-locus MIRU-VNTR. Differences in the VNTR analysis were restricted to a single-locus variant in locus MIRU 42 (424/Mtub04) (3 repetitions in the first isolate [VNTR variant A] and 1 repetition in the last isolate [VNTR variant B]) (**Figure 1A**). A mixture of VNTR variants A and B (detected by the observations of the corresponding double alleles in MIRU 42) was detected in the 2 intermediate specimens (August), thus leading us to select the first and last isolates for further genotyping.

According to RFLP, the differential IS6110 band detected in one of the 2 variants (**Figure 1A**) was localized using LM-PCR and mapped at the location described for one of the 16 insertions of H37Rv (3551227–3552587).

WGS revealed 11 different SNPs between the 2 variants. After comparing the WGS data with the most recent common ancestor of MTB as a reference, we obtained 6 specific SNPs from variant A and 5 from variant B (**Table 1**). Of the 6 specific SNPs from variant A, 3 were non-synonymous and 3 were synonymous. Of the 5 SNPs from variant B, 4 were non-synonymous, and the remaining one was intergenic.

Analysis of Functional Relevance

The next step in the characterization of these clonal variants was to determine the functional relevance of the subtle modifications identified. The IS6110 variation mapped in a hotspot for



**TABLE 1 | Specific SNPs from variants obtained by WGS.**

Variant A	Ancestor	S/NS/I	Gene
T	C	NS	Rv1963c ( <i>mce3R</i> )
C	A	NS	Rv2579 ( <i>dhaA</i> )
T	C	NS	Rv2921c ( <i>ftsY</i> )
T	C	s	Rv0836c
G	T	s	Rv1487
G	C	s	Rv1497 ( <i>lipL</i> )
Variant B	Ancestor	S/NS/I	Gene
T	C	NS	Rv1201c
T	G	NS	Rv1527c ( <i>pks5</i> )
A	G	NS	Rv1553 ( <i>frdB</i> )
G	C	NS	Rv2209
C	G	I	Rv2644c–Rv2645

S, synonymous SNP; NS, Non-synonymous SNP; I, Intergenic SNP.

the IS6110 insertion sequence, thus minimizing the potential functional significance of this variation. The MIRU-VNTR modifications mapped in a region between 2 stop codons for 2 adjacent genes (Rv0353 and Rv0354), again reducing the likelihood of a functional impact for these changes. In contrast, the fact that some of the SNP-based variability revealed by WGS corresponded to non-synonymous changes in relevant genes made it worthwhile to evaluate their functional impact.

We evaluated the potential impact of SNP-based variability by focusing on the SNP in Rv1963c (*mce3R*), because the SNP maps in a repressor of several genes included in the *yrbE3A*–Rv1971 operon (Santangelo et al., 2008; Figure 2). We compared the expression of the first gene in the operon, *yrbE3A*, in the 2 variants. A 0.7910-fold ratio (0.6936–0.8884;  $p < 0.05$ ) was observed in the expression of *yrbE3A* in variant A compared with variant B, indicating that the SNP detected in variant A increased the efficiency of the repressor Mce3R.

## Infectivity of Clonal Variants

The infectivity of the clonal variants was measured in both murine and cellular models. In the murine model, no significant differences were found in the growth rates of each variant in the lung ( $0.8039 \pm 0.1707$  and  $0.8658 \pm 0.1683$  for variants A and B) or in the spleen ( $0.2328 \pm 0.2996$  and  $0.5926 \pm 0.1911$  for variants A and B).

To evaluate the presence of subtle differences between the infectivity of the variants that were not revealed in standard infection assays, we performed competitive assays by simultaneously infecting mice with the 2 clonal variants. We did not detect differences in the proportion of the variants at day 1 compared with the proportion in the inoculum (both in lung and in spleen Figure 3A). However, the representativeness of variant B between day 1 and week 5 increased both in the lung (from  $0.3095 \pm 0.0336$  at day 1 to  $0.5058 \pm 0.0950$  at week 5) and in the spleen (from  $0.2019 \pm 0.0624$  at day 1 to  $0.3638 \pm 0.0779$  at week 5) ( $p < 0.05$ ) (Figure 3A).

In order to obtain new evidence of the likely higher infectivity of clonal variant B, we performed a new co-infection experiment. On this occasion, we used macrophages and modified the inoculum to force underrepresentation of clonal variant B (96:4). Even at this unbalanced proportion, more efficient uptake was observed for variant B than for variant A, meaning that the 96:4 proportion in the inoculum became 61:39 at 3 h (Figure 3B). Measurements could not be taken appropriately at day 7 owing to an unexpected finding, namely, the emergence of a previously undetected clonal variant (variant C: 2 repetitions at locus MIRU 42, Figure 1B). The 3 variants were now found at a proportion 62:10:28 (A:B:C). Consequently, the representativeness of variant A remained constant compared with that observed at 3 h after infection, whereas the presence of variant B decreased.

## Characterization of the New Clonal Variant

MIRU-VNTR analysis indicated that variant C only harbored a difference in locus MIRU 42 (allelic value 2). The IS6110-RFLP type was identical to that of clonal variant B, suggesting that variant B was its parental strain (Figure 1B). WGS of the new variant revealed no SNPs with respect to variant B, confirming that variant C derived from variant B.

The fitness of variant C was compared with that of the other 2 variants and proved to be higher, as indicated by its shorter lag phase ( $219 \pm 5.292$  h vs.  $260 \pm 12.12$  h [variant B] and  $255.3 \pm 12.5$  [variant A];  $p < 0.01$ ) (Figure 4A) and rate of growth ( $25.62 \pm 1.22$  vs.  $71.23 \pm 16.32$  h [variant B] and  $56.84 \pm 5.569$  h [variant A];  $p < 0.05$ ) (Figure 4B).

## Analysis of Inpatient Microevolution Dynamics

In order to determine the dynamics by which variant A (July 2000) was replaced by variant B (November 2001), we recovered the 2 intermediate specimens (August 2001, specimens 1 and 2) in which the 2 variants coexisted to evaluate whether it was possible to track the progressive elimination of variant A. Taking advantage of the differential SNPs found between the 2 variants, we selected the seven non-synonymous SNPs to assess the representativeness of each variant by analyzing multiple single colonies. For this aim we tailored allele specific PCRs to directly track the two alternative alleles in the seven SNPs. The ASO-PCR targeted to the analysis of the SNP in Rv2921 could not be optimized. The analysis of the remaining six SNPs on 30 colonies (12 and 18 from specimens 1 and 2, respectively) revealed that in the specimen 1 coexisted the variant A with two novel variants (new variant 1 and 2), sharing 4 and five alleles out of six, respectively, with variant B (Table 2). In the specimen 2 we observed the coexistence of variant A with the new variant 2.

The detailed analysis of the intermediate isolates allowed us to rule out our firstly assumed hypothesis of a chronological substitution of variant A by variant B. Instead, there was a coexistence of new evolutionary intermediates acquiring sequentially the variant B alleles (Figure 5A). In summary we could identify 4 independent clonal variants involved at different stages of the microevolutionary process along the infection.



FIGURE 2 | Diagram showing the location of the gene *mce3R* within the *mce3R* operon.

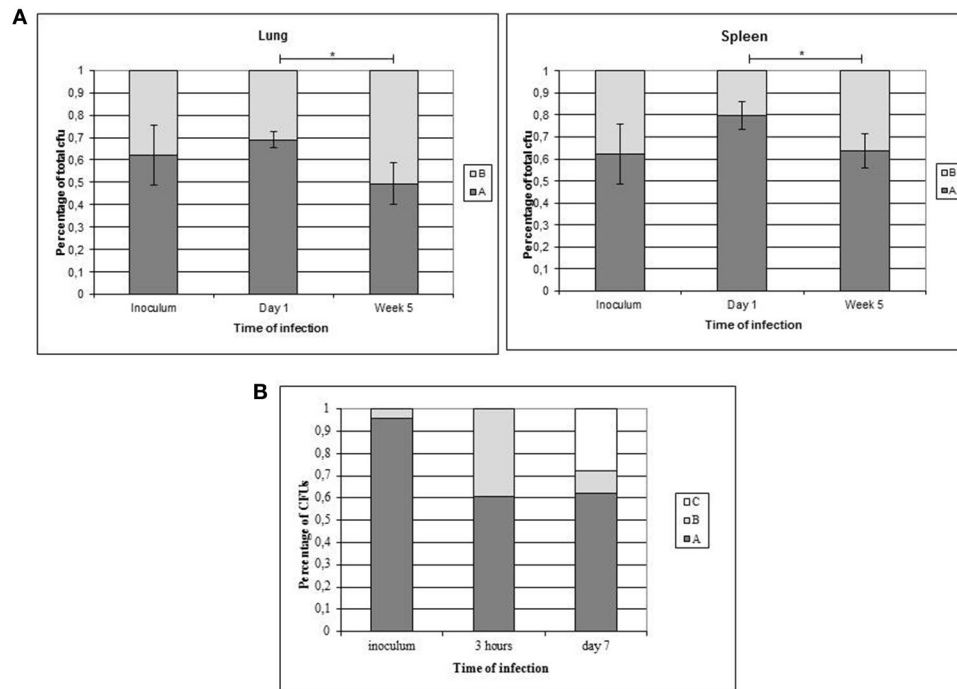


FIGURE 3 | Co-infection assays with both clonal variants. Asterisks indicate statistical significance ( $p < 0.05$ ). (A) Proportion of co-infecting variants obtained in the lung (left panel) and spleen (right panel) from 5 mice at each time point. (B) Co-infection of THP-1 cells with the clonal variants.

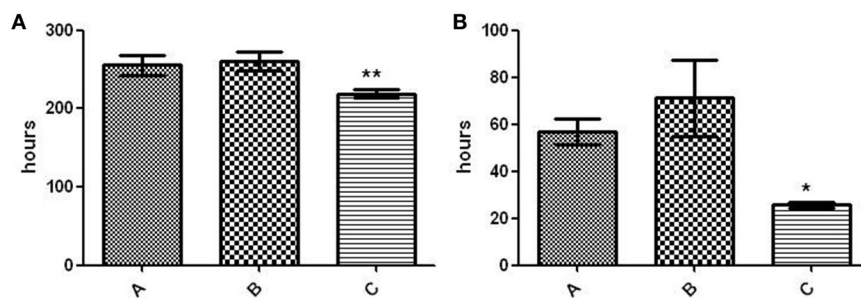


FIGURE 4 | Fitness assay of the 3 variants. (A) Lag phase (time to positive threshold [75 GU]). The asterisks indicate statistical significance ( $p < 0.01$ ). (B) Growth rate (time required for the 4,000–6,000-GU increase). The asterisk indicates statistical significance ( $p < 0.05$ ).

## Analysis of Interpatient Microevolution Dynamics in Clustered Patients

Our final step was to search for other cases in the population involved in a transmission cluster with the case we report. We searched for cases infected by each of the initial and final variants

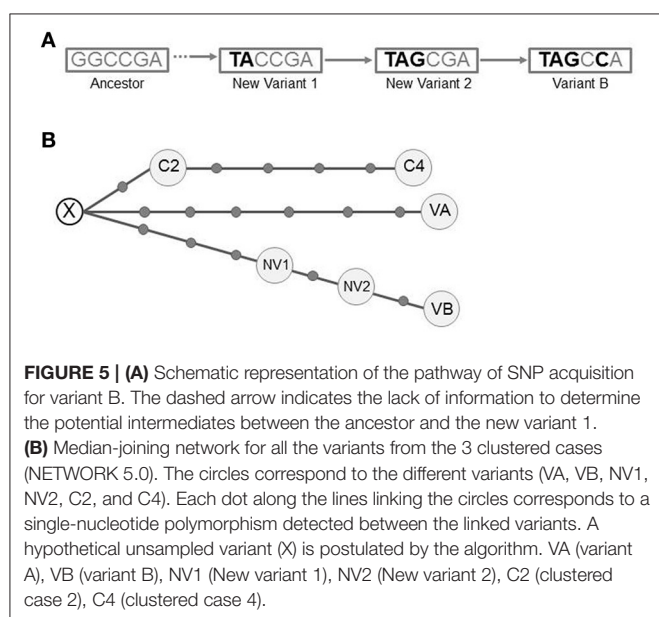
A and B. For variant A, we identified 2 cases (C2 and C4, years 2003 and 2004 respectively) and for variant B none.

We performed WGS analysis with these 2 new isolates. Neither shared the alleles found in the SNPs of variants A and B with respect to the ancestor, indicating that they had

**TABLE 2 |** Distribution of SNPs (in bold, after comparing to the ancestor reference) identified in the clonal variants by analyzing single colonies.

SNP:Gene	Ancestor	Episode 1		Episode 2				
		July 2000	August 2001:Specimen 1 (12 colonies)			August 2001:Specimen 2 (20 colonies)		November 2001
		Variant A	Variant A (3)	New variant 1 (2)	New variant 2 (7)	Variant A (6)	New variant 2 (14)	Variant B
SNP1:Rv1201	G	G	G	<b>T</b>	<b>T</b>	G	<b>T</b>	<b>T</b>
SNP2:Rv1527	G	G	G	<b>A</b>	<b>A</b>	G	<b>A</b>	<b>A</b>
SNP3:Rv1553	C	C	C	C	<b>G</b>	C	<b>G</b>	<b>G</b>
SNP4:Rv1963	C	<b>T</b>	<b>T</b>	C	C	<b>T</b>	C	C
SNP5:Rv2209	G	G	G	G	G	G	G	<b>C</b>
SNP6:Rv2579	A	<b>C</b>	<b>C</b>	A	A	<b>C</b>	A	A
SNP7:Rv2921	C	<b>T</b>						C

The number between brackets indicates the number of colonies in which that composition of SNPs was identified.



followed an independent microevolutionary branch (**Figure 5B**) from a common unsampled node (X). C2 and C4 were sequential steps of that branch. The SNP acquired in C2 was synonymous and mapped in Rv0037 (conserved membrane protein). Three of the 4 SNPs from C4 mapped in Rv0015 (non-synonymous, transmembrane serine/threonine-protein kinase A), Rv1821 (synonymous, preprotein translocase ATPase secA2), and Rv2934 (non-synonymous, phenolphthiocerol synthesis type-I polyketide synthase). The fourth SNP was intergenic.

## DISCUSSION

We performed an in-depth study of the clonal variants that emerged in a patient with recurrent tuberculosis. Initial genotypic characterization confirmed that the strains isolated from each episode were clonal variants because they only differed in a single-locus variant in 24 MIRU loci and in a single IS6110 copy. Similarly subtle differences have been reported for clonal variants

in various studies (Shamputa et al., 2006; Al-Hajj et al., 2010; Navarro et al., 2011).

WGS was applied to obtain a more exhaustive description of the degree of variability between the clonal variants, and 11 SNPs were identified. These figures are higher than expected according to the variability observed in transmission chains (similarity threshold of 12 SNPs before isolates can be considered clustered; Walker et al., 2013). Our findings indicate that intrapatient variability itself could be close to this threshold. Also, given previously reported mutation rates in MTB, usually within the range of 0.25–0.5 SNPs per genome per year, the accumulation of SNPs in the inpatient variants is much higher than expected. It suggests a higher tendency to acquire variability by this strain, but we must also leave open the possibility that part of that variability could have been acquired during the latency period, as it has been suggested elsewhere (Lillebaek et al., 2016).

We compared the alleles found in the 11 SNPs with those found in the most recent common ancestor of MTB as a reference (Comas et al., 2010, 2013) and found that variant B did not evolve directly from variant A, because 6 and 5 alleles found in the 11 SNPs were specific to variants A and B, respectively. Therefore, each variant represented an independent evolutionary path from a common parental strain that had not been sampled in the specimens analyzed. The results obtained from this new analysis fitted better with the consensus variability thresholds (Walker et al., 2013), thus illustrating that it is essential to integrate allelic data from the ancestor in order to interpret SNP data appropriately and that this step is needed to establish the true phylogenetic relationships between clonal variants.

Our naïve initial interpretation of 2 inpatient variants emerging one from the other became a more complex picture of 2 independent evolutionary branches from an unsampled common ancestor. This complexity increased after we found that the variants from the intermediate specimens (August), which were initially interpreted as co-infection of variants A and B, corresponded to mixtures of variant A with 2 novel intermediate variants. The detection of these new variants allowed us to infer partially the pathway followed in the emergence of variant B. Our findings are consistent with those reported for macaque models (Lin et al., 2014), where different clonal variants in different lesions evolved independently, that is, some finished in a *cul de*



*sac*, whereas others progressed. The finding of different clonal variants in different lesions could lead to a heterogeneous drain of variants into respiratory samples and differential identification of one variant or another.

The variability of the evolutionary routes explored by our strain was also illustrated by the analysis of a further 2 cases of tuberculosis in the population involved in the same transmission cluster as the first study patient. The isolates from these 2 cases shared the same evolutionary branch, although this was independent of that of the inpatient variants A and B.

Taken together, the SNP data and the subtle differences identified by RFLP and MIRU-VNTR suggest that the strain in question had a higher than expected tendency to microevolve. This is supported not only by its higher than expected accumulation of SNPs, but also by the variations according to IS6110 distribution and by the variations in the number of repetitions for certain VNTR loci. This tendency is not restricted to a specific host, because acquisition of variability was observed in all 3 clustered TB cases analyzed. The most obvious alert is that clusters involving strains of this kind are likely to exceed the similarity thresholds established to define clusters; therefore, related cases could be misinterpreted as unclustered.

The WGS analysis also provided clues about the potential functional significance of some of the above mentioned evolutionary routes followed by the clonal variants. In the case of specific SNPs for variant A, the most remarkable polymorphism is a non-synonymous substitution mapping in the essential gene *mce3R*, a transcriptional repressor of the *mce3R* regulon, which is involved in lipid metabolism and redox reactions (de la Paz Santangelo et al., 2009). The expression of the first gene in the *mce3* operon, *yrbE3A*, in variant A was lower than in variant B, indicating that the SNP found in variant A leads to higher efficiency of the repressor, which would likely have a functional effect. At some stage in the infection process, more marked repression of this operon might have been advantageous, since both variant A and one of the intermediate variants following a different evolutionary branch shared this SNP.

The *in silico* analysis of the SNPs detected in the clonal variants revealed the involvement of enzymatic, membrane, and cell division proteins. Together with differential expression for variants showing a SNP in *mce3R* and their differences in TNF $\alpha$  production when infecting macrophages (Martín et al., 2011), this involvement indicated the potential functional impact of several of the specific SNPs acquired for each clonal variant.

These findings led us to explore in detail the infective behavior of the clonal variants. Based on a competitive strategy when assaying infectivity in infection models developed in our group (Martín et al., 2011; Navarro et al., 2013) we revealed that clonal variant B was more successful than variant A. This advantageous behavior was observed even when underrepresented (Barczak et al., 2005), and, surprisingly, likely owing to the marked imbalance forced in the co-infection assay, a new variant with a higher fitness emerged, namely variant C, which had not been identified in the clinical specimens. This finding suggested a marked tendency of this strain to microevolve.

In summary, we present an in-depth study of multiple clonal variants which emerged in sequential stages in a patient with

recurrent tuberculosis through several independent exploratory branches in the same microevolution event. In addition, other microevolutionary branch leading to different variants was detected in other hosts sharing the same transmission cluster, and finally a novel variant not sampled from the clinical specimens emerged in an infection assay in the laboratory. We also observed differential gene expression and differential infectivity between some of the emerged variants.

These observations emphasize how complex and functional the microevolutionary phenomena in the infection by *M. tuberculosis* can be and indicate that some strains are especially prone to microevolution, which could impact on the inference of clusters based on WGS if strict thresholds are applied.

## ETHICS STATEMENT

The Institutional Animal Care and Use Committee (IACUC) of the Gregorio Marañón General Hospital (ES 280790000087) (Madrid, Spain) reviewed and approved the experimental protocol. The procedures followed were in agreement with the current Spanish Legislation (RD 53/2013), the European Directive 2010/63/UE [which follows the guidelines and recommendations approved by the Federation of Laboratory Animal Science Associations (FELASA)] and the ethical rules which are applied in this center.

## AUTHOR CONTRIBUTIONS

DG, LP-L, JS, IC, MH, YN, and OS: Made substantial contributions to the concept and design of the work through acquisition, analysis and interpretation of data. YN, LP-L, and DG: Drafted the work and provided critical revision for important intellectual content. All: Approved the final version of the article.

## ACKNOWLEDGMENTS

We are grateful to Thomas O'Boyle for proofreading the text. This study was funded by the Ministry of Economy and Competitiveness ISCIII-FIS (PI15/01554) and co-funded by ERDF (FEDER) funds from the European Commission, "A way of making Europe." Research by YN was supported by a PICATA pre-doctoral fellowship (BE55/11) from the Moncloa Campus of International Excellence (UCM-UPM, Instituto de Investigación Sanitaria Gregorio Marañón). LP-L was supported by a Miguel Servet contract (MS15/00075-CP15/00075). IC lab is financed by Ministerio de Economía y Competitividad (Spanish Government) research grant SAF2013-43521-R, SAF2016-77346-R, and the European Research Council (ERC) (638553-TB-ACCELERATE).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00694/full#supplementary-material>

## REFERENCES

- Akhtar, P., Singh, S., Bifani, P., Kaur, S., Srivastava, B. S., and Srivastava, R. (2009). Variable-number tandem repeat 3690 polymorphism in Indian clinical isolates of *Mycobacterium tuberculosis* and its influence on transcription. *J. Med. Microbiol.* 58, 798–805. doi: 10.1099/jmm.0.002550-0
- Al-Hajj, S. A., Akkerman, O., Parwati, I., al-Gamdi, S., Rahim, Z., van Soolingen, D., et al. (2010). Microevolution of *Mycobacterium tuberculosis* in a tuberculosis patient. *J. Clin. Microbiol.* 48, 3813–3816. doi: 10.1128/JCM.00556-10
- Alonso, M., Alonso Rodriguez, N., Garzelli, C., Garzelli, C., Martínez Lirola, M., Herranz, M., Samper, S., et al. (2010). Characterization of *Mycobacterium tuberculosis* Beijing isolates from the Mediterranean area. *BMC Microbiol.* 10:151. doi: 10.1186/1471-2180-10-151
- Barczak, A. K., Domenech, P., Boshoff, H. I., Reed, M. B., Manca, C., Kaplan, G., et al. (2005). *In vivo* phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J. Infect. Dis.* 192, 600–606. doi: 10.1086/432006
- Comas, I., Chakravarti, J., Small, P. M., et al. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* 42, 498–503. doi: 10.1038/ng.590
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45, 1176–1182. doi: 10.1038/ng.2744
- de la Paz Santangelo, M., Klepp, L., Nuñez-García, J., Blanco, F. C., Soria, M., García-Pelayo, M. C., et al. (2009). Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in *Mycobacterium tuberculosis*. *Microbiology* 155, 2245–2255. doi: 10.1099/mic.0.027086-0
- García de Viedma, D., Alonso Rodríguez, N., Andrés, S., Ruiz Serrano, M. J., and Bouza, E. (2005). Characterization of clonal complexity in tuberculosis by mycobacterial interspersed repetitive unit-variable-number tandem repeat typing. *J. Clin. Microbiol.* 43, 5660–5664. doi: 10.1128/JCM.43.11.5660-5664.2005
- Huyen, M. N., Kremer, K., Lan, N. T., Cobelens, F. G., Buu, T. N., Dung, N. H., et al. (2012). Mixed tuberculosis infections in rural South Vietnam. *J. Clin. Microbiol.* 50, 1586–1592. doi: 10.1128/JCM.00434-12
- Lillebaek, T., Norman, A., Rasmussen, E. M., Marvig, R. L., Folkvardsen, D. B., Andersen, Å. B., et al. (2016). Substantial molecular evolution and mutation rates in prolonged latent *Mycobacterium tuberculosis* infection in humans. *Int. J. Med. Microbiol.* 306, 580–585. doi: 10.1016/j.ijmm.2016.05.017
- Lin, P. L., Ford, C. B., Coleman, M. T., Myers, A. J., Gawande, R., Ioerger, T., et al. (2014). Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat. Med.* 20, 75–79. doi: 10.1038/nm.3412
- Martín, A., Herranz, M., Navarro, Y., Lasarte, S., Ruiz Serrano, M. J., Bouza, E., et al. (2011). Evaluation of the inaccurate assignment of mixed infections by *Mycobacterium tuberculosis* as exogenous reinfection and analysis of the potential role of bacterial factors in reinfection. *J. Clin. Microbiol.* 49, 1331–1338. doi: 10.1128/JCM.02519-10
- Muwonge, A., Kankya, C., Olea-Popelka, F., Biffa, D., Ssengooba, W., Berit, D., et al. (2013). Molecular investigation of multiple strain infections in patients with tuberculosis in Mubende district, Uganda. *Infect. Genet. Evol.* 17, 16–22. doi: 10.1016/j.meegid.2013.03.039
- Navarro, Y., Herranz, M., Pérez-Lago, L., Martínez Lirola, M., INDAL-TB, Ruiz-Serrano, M. J., et al. (2011). Systematic survey of clonal complexity in tuberculosis at a populational level and detailed characterization of the isolates involved. *J. Clin. Microbiol.* 49, 4131–4137. doi: 10.1128/JCM.05203-11
- Navarro, Y., Pérez-Lago, L., Sislera, F., Herranz, M., de Juan, L., Bouza, E., et al. (2013). Unmasking subtle differences in the infectivity of microevolved *Mycobacterium tuberculosis* variants coinfecting the same patient. *Int. J. Med. Microbiol.* 303, 693–696. doi: 10.1016/j.ijmm.2013.10.002
- Oelemann, M. C., Diel, R., Vatin, V., Haas, W., Rüsch-Gerdes, S., Loch, C., et al. (2007). Assessment of an optimized mycobacterial interspersed repetitive unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J. Clin. Microbiol.* 45, 691–697. doi: 10.1128/JCM.01393-06
- Olsen, I., Balasingham, S. V., Davidsen, T., Debebe, E., Rødland, E. A., van Soolingen, D., et al. (2009). Characterization of the major formamidopyrimidine-DNA glycosylase homolog in *Mycobacterium tuberculosis* and its linkage to variable tandem repeats. *FEMS Immunol. Med. Microbiol.* 56, 151–161. doi: 10.1111/j.1574-695X.2009.00562.x
- Pérez-Lago, L., Comas, I., Navarro, Y., González-Candelas, F., Herranz, M., Bouza, E., et al. (2014). Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J. Infect. Dis.* 209, 98–108. doi: 10.1093/infdis/jit439
- Pérez-Lago, L., Herranz, M., Lirola, M. M., Group, I.-T., Bouza, E., and García de Viedma, D. (2011). Characterization of microevolution events in *Mycobacterium tuberculosis* strains involved in recent transmission clusters. *J. Clin. Microbiol.* 49, 3771–3776. doi: 10.1128/JCM.01285-11
- Pérez-Lago, L., Navarro, Y., Herranz, M., Bouza, E., and García-de-Viedma, D. (2013). Differences in gene expression between clonal variants of *Mycobacterium tuberculosis* emerging as a result of microevolution. *Int. J. Med. Microbiol.* 303, 674–677. doi: 10.1016/j.ijmm.2013.09.010
- Santangelo, M. P., Blanco, F. C., Bianco, M. V., Klepp, L. I., Zabal, O., Cataldi, A. A., et al. (2008). Study of the role of Mce3R on the transcription of mce genes of *Mycobacterium tuberculosis*. *BMC Microbiol.* 8:38. doi: 10.1186/1471-2180-8-38
- Shamputa, I. C., Jugheli, L., Sadradze, N., Willery, E., Portaels, F., Supply, P., et al. (2006). Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. *Respir. Res.* 7:99. doi: 10.1186/1465-9921-7-99
- Tantivitayakul, P., Panapruksachai, S., Billamas, P., and Palittapongarnpim, P. (2010). Variable number of tandem repeat sequences act as regulatory elements in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 90, 311–318. doi: 10.1016/j.tube.2010.08.003
- van Embden, J. D., Cave, M. D., Crawford, J. T., Dale, J. W., Eisenach, K. D., Gicquel, B., et al. (1993). Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* 31, 406–409.
- Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., et al. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* 13, 137–146. doi: 10.1016/S1473-3099(12)70277-3
- Yindeeyoungyeon, W., Likitvatanavong, S., and Palittapongarnpim, P. (2009). Characterization of alpha-isopropylmalate synthases containing different copy numbers of tandem repeats in *Mycobacterium tuberculosis*. *BMC Microbiol.* 9:122. doi: 10.1186/1471-2180-9-122

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Navarro, Pérez-Lago, Herranz, Sierra, Comas, Sicilia, Bouza and García de Viedma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Double-Face Meets the Bacterial World: The Opportunistic Pathogen *Stenotrophomonas maltophilia*

Felipe Lira<sup>1</sup>, Gabriele Berg<sup>2</sup> and José L. Martínez<sup>1\*</sup>

<sup>1</sup> Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Universiti Sultan Zainal Abidin,  
Malaysia

### Reviewed by:

Radoslaw Pluta,  
International Institute of Molecular  
and Cell Biology in Warsaw (IIMCB),  
Poland  
Prabhu B. Patil,  
Institute of Microbial Technology  
(CSIR), India  
Gregory Anderson,  
Indiana University – Purdue University  
Indianapolis, United States

### \*Correspondence:

José L. Martínez  
jlmartnez@cnb.csic.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 31 July 2017

**Accepted:** 25 October 2017

**Published:** 09 November 2017

### Citation:

Lira F, Berg G and Martínez JL  
(2017) Double-Face Meets  
the Bacterial World: The Opportunistic  
Pathogen *Stenotrophomonas*  
*maltophilia*. *Front. Microbiol.* 8:2190.  
doi: 10.3389/fmicb.2017.02190

Most studies on bacterial virulence focus on the pathogen itself. However, it is important to recall that the in-host behavior and the virulence of bacterial pathogens constitute a complex situation that depends on both the microorganisms and the infected host. While healthy people (the community) is infected by classical pathogenic microorganisms, able to cope with the anti-infection defenses of the host, in the case of people with basal diseases, debilitated or immunodepressed, the range of pathogens able to cause infection is wider and includes the so-named opportunistic pathogens, which lack the inherent ability to cause disease in healthy hosts and rarely produce infections in the community. Some of the most relevant opportunistic pathogens, as *Stenotrophomonas maltophilia*, have an environmental origin and, in occasions, present interesting biotechnological properties. Consequently, it is important knowing whether *S. maltophilia* isolates recovered from infections constitute a specific phylogenetic branch that has evolved toward acquiring a virulent phenotype as it happens in the case of classical pathogens or rather, any member of this bacterial species is capable of producing infection and its pathogenic behavior is mainly a consequence of the host situation. To address this question, we analyzed a set of environmental and clinical *S. maltophilia* strains. Our results indicate that this opportunistic pathogen presents a large core genome and that the distribution of genes in general, and of known virulence determinants in particular, is similar among environmental and clinical isolates. The majority of genes not belonging to the *S. maltophilia* core genome are present in just one or two of the analyzed strains. This indicates that, more than speciation into different lineages (virulent and environmental), the evolution of *S. maltophilia* is based in the strain-specific acquisition of genes, likely involved in the adaptation of this bacterial species to different microniches. In addition, both environmental and clinical isolates present low susceptibility to several antimicrobials. Altogether our results support that *S. maltophilia* does not present a specific evolutionary branch toward virulence and most likely infection is mainly the consequence of the impaired anti-infective response of the infected patients.

**Keywords:** *Stenotrophomonas maltophilia*, opportunistic pathogens, comparative genomics, pangenome, core genome, antibiotic resistance

## INTRODUCTION

*Stenotrophomonas maltophilia* is an opportunistic pathogen, with an environmental origin, which causes a variety of infections at hospitals (Brooke, 2012; Chang et al., 2015; Jeon et al., 2016; Brooke et al., 2017), particularly in those patients under previous therapy with broad-spectrum antibiotics (Chang et al., 2015), and in patients with underlying diseases as cystic fibrosis (Nicoletti et al., 2011; Pompilio et al., 2016; Esposito et al., 2017). *S. maltophilia* infections are difficult to treat because this pathogen displays low susceptibility to several antimicrobials (Sanchez et al., 2009; Sanchez, 2015). As the consequence of this situation and likely also because *S. maltophilia* infects mainly severely debilitated individuals, the mortality of patients suffering *S. maltophilia* infections is high (Jeon et al., 2016). Consequently, understanding the underlying features by which this pathogen can traverse different ecological allocations, from its natural habitat toward infecting humans, may help in the development of strategies to improve the treatment of infections due to this microorganism.

Besides its clinical relevance, different *S. maltophilia* strains exert an extraordinary range of activities with biotechnological relevance (Mukherjee and Roy, 2016), such as bioremediation (Dungan et al., 2003; Berg and Martinez, 2015), degradation of toxic compounds (Lee et al., 2002), biosynthesis (Jakobi et al., 1996; Nangia et al., 2009; Zonaro et al., 2015) and biological control in agriculture (Dunne et al., 2000; Alavi et al., 2013), among others.

Given these two aspects of *S. maltophilia*, it is highly relevant to determine whether infective and environmental (non-clinical) *S. maltophilia* isolates constitute different evolutionary branches in this species as it has been shown in the case of the *Burkholderia cepacia* complex (Chiarini et al., 2006) or if, by contrary, any strain can infect a compromised human host, as it has been described for *Pseudomonas aeruginosa* (Alonso et al., 1999; Morales et al., 2004; Wiehlmann et al., 2007). This is particularly relevant in order to evaluate the risks for human health associated to the use of *S. maltophilia* for biotechnological purposes, mainly for non-confined applications, as agriculture.

Different works, based in classical Multi-Locus Sequence Typing (MLST), *in silico* MLST and whole genome analyses, have been published to address the phylogenetic structure of this species and of others belonging to the same complex (Rocco et al., 2009; Adamek et al., 2014; Gherardi et al., 2015; Youenou et al., 2015; Esposito et al., 2017; Ochoa-Sanchez and Vinuesa, 2017). Nevertheless, it is still unclear whether or not clinical isolates are predominant in any of these branches. In addition, studies on the potential correlation between the presence in the genome of virulence determinants and antibiotic resistance with the origin of the strains (clinical or environmental) are extremely limited, despite the relevance of these features for the nosocomial infections produced by *S. maltophilia*.

In order to address whether or not clinical and environmental isolates belong to different phylogenetic branches in *S. maltophilia*, in the present work we have sequenced 20 *S. maltophilia* isolates (10 from clinical environments and

10 from environmental samples). Four complete genomes sequences were also included in the study as references, two clinical strains *S. maltophilia* K279a (Crossman et al., 2008) and D457 (Lira et al., 2012) and two environmental isolates, *S. maltophilia* R551-3 (Lucas et al., 2008) and JV3 (Lucas et al., 2011). In addition, in the present work we present the phenotypic analysis of the studied isolates in order to determine whether or not clinical isolates are more resistant to antibiotics than environmental ones, information that cannot be obtained from the simple inspection of *S. maltophilia* available genomes.

## MATERIALS AND METHODS

### DNA Extraction and Genome Sequencing of 20 New Strains of *Stenotrophomonas maltophilia*

The complete DNAs of 20 isolates of *S. maltophilia* (Table 1) were extracted using the GENOME DNA Kit (MP Biomedicals LLC, Illkrich, France). Whole-genome sequencing was performed at the facility of the Madrid Science Park (Madrid, Spain), using Illumina MiSeq technology (Illumina, San Diego, CA, United States) from DNA libraries with insertion sizes between 700 and 800 bp, to generate paired-end reads with 260–300 bp length.

### Quality Control and Sequence Assembling

Quality score of the sequences of all strains was checked using FastQC v.0.11.2, to identify adapters and contaminant sequences remaining after sequencing. Contaminant sequences were removed using the AlienTrimmer v.0.4.0 software (Crisciuolo and Brisse, 2013) and a customized database of adapters adding the contaminant sequences recognized by FastQC. Sequence trimming and filtering were performed by PRINSEQ-Lite (Schmieder and Edwards, 2011) to filter the sequences by length and quality score (Phred  $\geq$  22, minimum read length = 90 bp). Each set of reads was submitted to *de novo* assembling using the Spades v.3.9 assembler (Bankevich et al., 2012) in a local server (24 cores and 512Gb RAM). After assembling, contigs with a minimal 5,000 bp length were selected. The synteny of the generated contigs was ordered using Mauve aligner (Darling et al., 2004) and two reference genomes, the model strains *S. maltophilia* D457 (Lira et al., 2012) and *S. maltophilia* K279 (Crossman et al., 2008). Both genomes were chosen because they were the largest complete genomes available. Contigs alignment did not presented divergences with respect to the reference genomes synteny.

### Open Reading Frames Detection, Gene Prediction and Annotation

For the prediction and annotation of the Open Read Frames (ORFs) from each set of contigs, we used two approaches: (a) In a first step the ORFs were predicted using Prodigal v2.6.1



(Hyatt et al., 2010), avoiding truncated genes. The parameters were set to predict genes containing both start and stop codons. This approach allowed the elimination of fragmented genes located at the edges of the contigs. Predicted ORFs were annotated performing a local alignment with BLASTp (Camacho et al., 2009) against the NCBI non-redundant database setting the expected value (*e-value*) of  $10^{-10}$ . In a second step, all contigs were submitted to the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Divergences between the local annotation and the PAGP were checked and curated manually.

## Comparative Genomics

Twenty draft genomes of *S. maltophilia* obtained in this study, and the complete genomes of four strains were used to estimate the preliminary core genome and pangenome sizes of *S. maltophilia*. The complete genomes of two clinical strains: D457 (NC\_017671.1) and K279a (NC\_010943.1) and of two environmental strains: R551-3 (NC\_011071.1) and JV3 (NC\_015947.1) were also used for the analysis. The accession numbers of draft genomes of the 20 strains of *S. maltophilia* analyzed in this study are: clinical strains: E729 (NERH00000000), E759 (NERG00000000), E999 (NERF00000000), G51 (NERE00000000), E301 (NERD00000000), D388 (NERC00000000), E861 (NERB00000000), C357 (NERA00000000), E539 (NEQZ00000000), E824 (NEQY00000000); environmental strains: NS26 (NEQO00000000), EP13 (NEQX00000000), EA22 (NEQW00000000), EA1 (NEQV00000000), PS5 (NEQU00000000), EA23 (NEQT00000000), EP20 (NEQS00000000), EP5 (NEQR00000000), EA21 (NEQQ00000000), EA63 (NEQP00000000) (Table 1).

The pangenome and the core genome of the sequenced strains were analyzed using the script GET\_HOMOLOGUES v.07112016 (Contreras-Moreira and Vinuesa, 2013). Clusters of homologous gene families were generated using the COGtriangles algorithm. To form clusters and estimate the core genome and pangenome of *S. maltophilia*, coverage and identity thresholds of 90% and of 95%, respectively were used.

The complete Coding DNA Sequence (CDS) composition and the clusters generated for all strains were used to perform a comparative analysis and to calculate the genome similarity distance to determine the relationship of clinical and environmental isolates. Clustered genes were used to compile the corresponding pangenome matrix using the script compare\_cluster.pl with default settings, embedded in the GET\_HOMOLOGUES software package. The clusters formed were classified considering the distribution of ortholog genes through the strains. The core genome contains those genes belonging to all strains, the soft-core genome the genes present in, at least, 95% of the strains, the shell genome the genes present in less than 95% and more than 10% of the genomes and the cloud genome the genes present in less than 10% of the genomes (Koonin and Wolf, 2008; Kaas et al., 2012).

**TABLE 1** | Accession numbers of the genomes and origins of the *S. maltophilia* isolates of used in this study.

	Strains	Accession number	Origin	Reference
Clinical strains	E729	NERH00000000	Urine	Alonso and Martinez, 2001
	E759	NERG00000000	Sputum	Alonso and Martinez, 2001
	E999	NERF00000000	Respiratory secretion	Alonso and Martinez, 2001
	G51	NERE00000000	Blood	Alonso and Martinez, 2001
	E301	NERD00000000	Urine	Alonso and Martinez, 2001
	D388	NERC00000000	Urine	Alonso and Martinez, 2001
	E861	NERB00000000	Sputum	Alonso and Martinez, 2001
	C357	NERA00000000	Urine	Alonso and Martinez, 2001
	E539	NEQZ00000000	Pus from a wound	Alonso and Martinez, 2001
	E824	NEQY00000000	Blood	Alonso and Martinez, 2001
	K279a*	NC_010943.1	Blood	Crossman et al., 2008
	D457*	NC_017671.1	Respiratory secretion	Lira et al., 2012
Environmental strains	NS26	NEQO00000000	Dune soil	De Boer et al., 2001; Ribbeck-Busch et al., 2005
	EP13	NEQX00000000	Rhizosphere of oilseed rape	Minkwitz and Berg, 2001
	EA22	NEQW00000000	Sewage	Minkwitz and Berg, 2001
	EA1	NEQV00000000	Brackish water	Minkwitz and Berg, 2001
	PS5	NEQU00000000	Rhizosphere of oilseed rape.	Berg et al., 1996
	EA23	NEQT00000000	Eye-care solution	Suckstorf and Berg, 2003
	EP20	NEQS00000000	Rhizosphere of potato	Minkwitz and Berg, 2001
	EP5	NEQR00000000	Rhizosphere of <i>Brassica napus</i> L.	Minkwitz and Berg, 2001
	EA21	NEQQ00000000	Sewage	Minkwitz and Berg, 2001
	EA63	NEQP00000000	Sewage	Gabrielle Berg's lab collection
	R551-3*	NC_011071.1	Endosphere	Lucas et al., 2008
	JV3*	NC_015947.1	Rhizosphere	Lucas et al., 2011

\*Complete genomes of *S. maltophilia*.

## In Silico Multi-Locus Sequences Typing and Polymorphic Sites in the Core Genome

*In silico* MLST analysis (Larsen et al., 2012) was performed using the web server of the Centre for Genomic Epidemiology<sup>1</sup>. The alleles from each strain were identified individually and their nucleotides sequences were further concatenated (separated by 10 Ns) to perform a Multiple Sequence Alignment (MSA) using ClustalW2. A phylogenetic tree based in this alignment was calculated using the same software based on the similarity distance between concatenated sequences.

The identification of polymorphic sites was performed using Snippy<sup>2</sup> using *S. maltophilia* K279a as reference strain (Accession number: NC\_010943). Polymorphic sites in genes shared by all strains formed the core of Single Nucleotide Polymorphisms SNPs, that was used to perform a MSA. A phylogenetic tree from the derived information was constructed by using the maximum likelihood method.

## Genomic Composition and Comparative Genomics

Putative functional similarities and differences between the clinical and the environmental strains were estimated by a

subsystem classification using the RAST server<sup>3</sup> (Aziz et al., 2008) and the coding sequences from each genome were classified according to their protein families (*FIGfams*). All strains were compared by the presence/absence of 20 subsystems and 35 functional roles included in the category of “Virulence, Disease and Defense”. A local database containing a set of specific genes, described as responsible for the virulence phenotype of *S. maltophilia* (Adamek et al., 2014) was used to retrieve similar genes from the studied strains. Hierarchical clustering was performed in R functions (Langfelder and Horvath, 2012). For this purpose, each one of the resulting tables containing the information about presence/absence of these genes was converted into a square similarity matrix to measure the distance between strains (R function ‘dist’), clustered based on the matrix data (R function ‘hclust’) and plotted as heatmap (R function ‘heatmap.2’).

## Quorum-Sensing Signals

It has been described that the alleles of the quorum-sensing system (QS) *rpfF* gene, *rpfF1* (GenBank: KJ149475) and *rpfF2* (GenBank: KJ149552), are markers of two different phylogenetic branches, each one presenting differences in terms of virulence (Huedo et al., 2014). To address whether or not the presence of a specific *rpfF* allele could be linked to clinical strains, the 108 N-terminal residues of RpfF, which has been proposed

<sup>1</sup>www.cbs.dtu.dk/services/MLST

<sup>2</sup>https://github.com/tseemann/snippy

<sup>3</sup>http://rast.nmpdr.org/

**TABLE 2 |** Overall characteristics of the genomes analyzed in the current article.

	Strains	Bases	Contigs	Largest contig	GC%	Predicted genes
Clinical strains	K279a*	4,851,126	1	4851126	66.3	4354
	D457*	4,769,156	1	4769156	66.8	4254
	E861	4,658,203	31	653242	66.4	4191
	D388	4,659,986	30	740754	66.4	4190
	E539	4,555,541	18	1731480	66.5	4057
	C357	4,810,581	17	954550	66.2	4310
	E824	5,041,912	14	1834293	65.9	4502
	E729	5,005,550	12	1548184	66.6	4540
	E999	4,414,069	11	1140634	66.7	3879
	G51	4,852,740	8	2066793	66.1	4368
	E301	4,428,328	5	3885998	66.8	3965
	E759	4,546,405	4	2470865	66.5	4083
Environmental strains	R551-3*	4,573,969	1	4573969	66.3	4023
	JV3*	4,544,477	1	4544477	66.9	4040
	EA23	4,752,304	29	642831	66.4	4283
	EP13	4,755,757	27	744273	66.4	4281
	NS26	4,689,165	18	1729723	66.2	4152
	EP20	4,625,290	16	2060034	66.1	4087
	EA1	4,752,176	16	918730	66.6	4234
	EA22	4,759,594	10	1721891	66.2	4265
	EA63	4,885,042	10	1847362	66	4390
	EA21	4,732,256	9	1707011	66.2	4246
	PS5	4,600,476	7	2135136	66.4	4076
	EP5	4,600,182	7	2134905	66.4	4075

\*Complete genomes of *S. maltophilia*.

to be used as markers for distinguishing the two RpfF variants (Huedo et al., 2014) were aligned using ClustalW2 (Larkin et al., 2007). A phylogenetic tree derived from this information was established using JalView v.2 (Waterhouse et al., 2009)

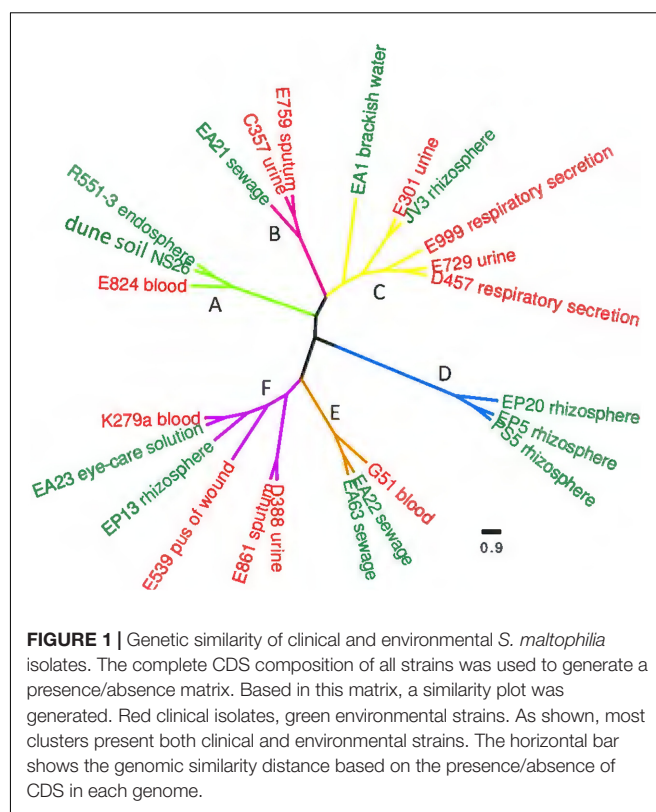
## Antibiotic Susceptibility

Minimum Inhibitory Concentrations (MICs) were determined in Mueller Hinton agar medium using MIC Test strips (Liofilchem) of the following antibiotics Trimethoprim/Sulfamethoxazole (SXT); Tigecyclin (TGC); Ceftazidime (CAZ); Cefepime (PM); Gentamicin (CN); Gatifloxacin (GAT); Colistin (CS); Chloramphenicol (CL); Imipenem (IMI); Ertapenem (ETP); Moxifloxacin (MXF); Nalidixic Acid (NA).

## RESULTS AND DISCUSSION

### Genome Assembling and Annotation of Clinical and Environmental Strains of *Stenotrophomonas maltophilia*

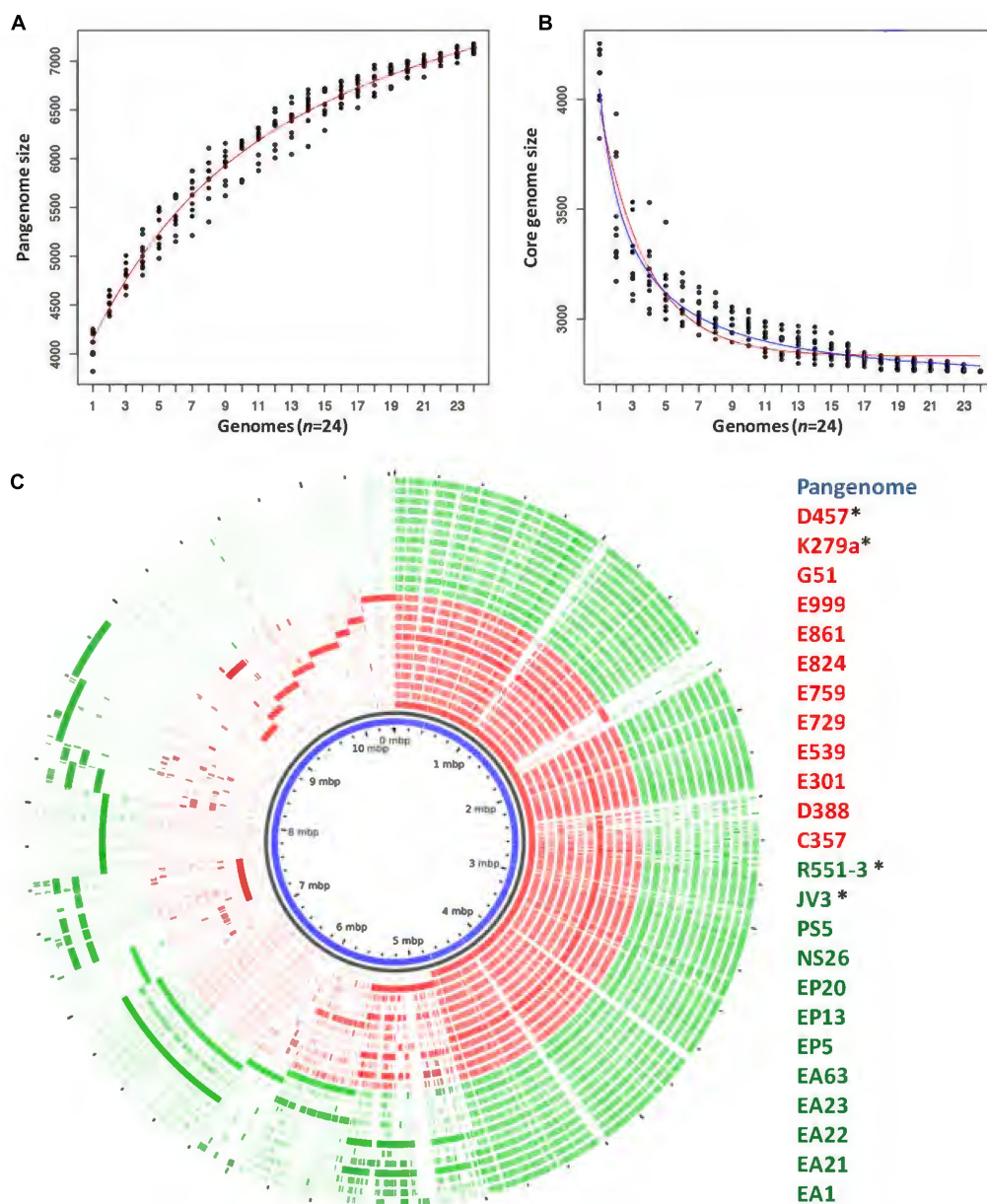
Although the number of sequenced genomes of the opportunistic pathogen *S. maltophilia* has increased since the first genome was published, specific analyses on the core genome and pangenome (Esposito et al., 2017) as well as on the genomic relationships of clinical and environmental isolates of this species are scarce. In addition, the quality (in terms of number of contigs) of the different available draft genomes is diverse, which makes their comparison difficult in occasions. Finally, clear information on the origin of the isolates (clinical or environmental) is not always available, making the use of these sequences difficult for the purposes of this work. Consequently, to analyze whether clinical and environmental isolates present different genomic features or, by contrary they do not form two different phylogenetic branches, we decided to sequence and analyze twenty isolates of *S. maltophilia* for which the origin has been well established (10 clinical and ten environmental). The assembling of all strains generated a total of 94 Mbp comprising 299 contigs. The genome length average of the sequenced strains was 4.7 Mb and their average GC% content 66.36% (Table 2). These data were similar to those of the available *S. maltophilia* complete genomes from strains D457, K279a, R551-3 and JV3, whose genome length and GC% content are, in average, 4.6 Mb and 66.57%, respectively. All contigs were submitted to the Prokaryotes Genome Annotation Pipeline (PGAP) (Tatusova et al., 2014) from NCBI, retrieving an average of 4206 CDS/strain (min = 3879; max = 4540) (Table 2). A presence/absence matrix was generated and used for the phylogenetic clustering of the different isolates based in the CDS composition of their genomes. As shown in Figure 1, and although branch D comprised just strains isolated from the rhizosphere, the other branches included both clinical and environmental isolates. This fact indicates that, at least in a whole view, there is not a clear differentiation in the CDS composition between the genomes of clinical and environmental *S. maltophilia* isolates.



**FIGURE 1 |** Genetic similarity of clinical and environmental *S. maltophilia* isolates. The complete CDS composition of all strains was used to generate a presence/absence matrix. Based in this matrix, a similarity plot was generated. Red clinical isolates, green environmental strains. As shown, most clusters present both clinical and environmental strains. The horizontal bar shows the genomic similarity distance based on the presence/absence of CDS in each genome.

### Effect of the Origin of *S. maltophilia* Isolates in Their Pangenome and Core Genome

The pangenome and the core genome of *S. maltophilia* were calculated using the draft genomes of the 20 sequenced strains as well as the four full genomes used as references in our work. The number of total genes was plotted as a function of the number of genomes added to the analysis. As shown in Figure 2A, an asymptotical increase in the number of genes with respect to the number of analyzed strains was detected. In agreement with previous information (Yu et al., 2016), this indicates that *S. maltophilia* has an open pangenome based on the analysis of the 24 genomes examined. To estimate the core genome, the number of genes shared by all stains was plotted as a function of the number of *S. maltophilia* genomes sequentially added to the analysis (Figure 2B). The core genome was estimated in 2762 genes, corresponding to 38% of the pangenome of *S. maltophilia* (Figure 2B). To estimate the tendency of the core genome two approaches were performed. Following the approach and terminology of Tettelin et al. (2005), the *S. maltophilia* core genome presents a 'relative constancy' after several genomes are included in the analysis (red line in Figure 2B), whereas the predictions using the approach of Willenbrock et al. (2007) is that the incorporation of novel genomes should produce a decay in the number of genes that compose the core genome of *S. maltophilia* (blue line in Figure 2B). Compositional analysis retrieved a pangenome composed by 7108 orthologous groups, although this number should likely increase when more genomes

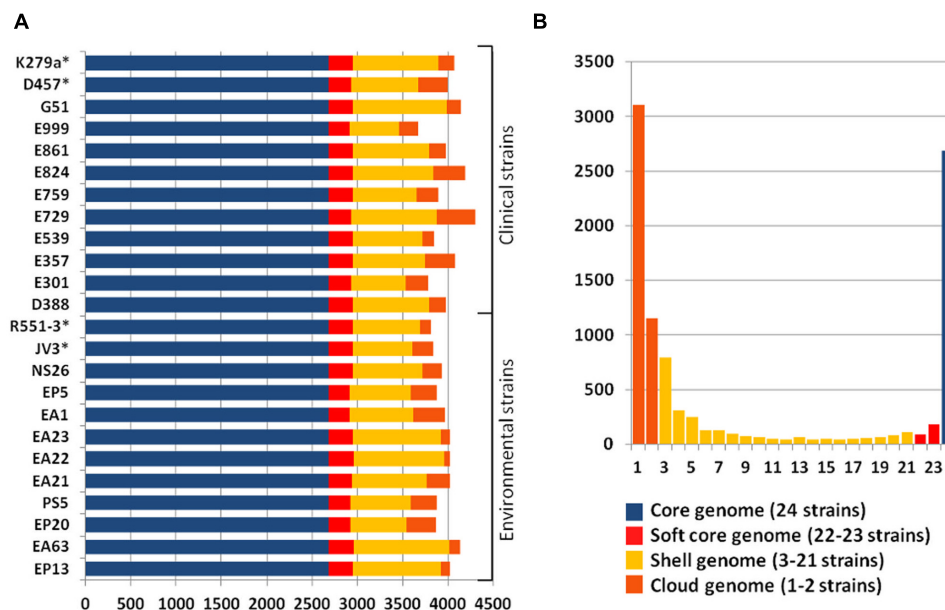


**FIGURE 2 |** Comparison of the genomes of environmental and clinical *S. maltophilia* isolates. The pangenome and the core genome of *S. maltophilia* were calculated by random sampling of the 24 genomes: **(A)** The number of total genes plotted as a function of the number of genomes added to the analysis is presented. As shown, *S. maltophilia* has an open pangenome. **(B)** The curve shows the number of genes shared by all strains as a function of the number of genomes of *S. maltophilia* added sequentially. Red and blue lines were plotted as an estimation of the tendency of the core genome. Red line indicates that the core genome of *S. maltophilia* should maintain, following the terminology and the estimation rules of Tettelin et al. (2005), a 'relative constancy' after several genomes are included in the analysis. Blue line indicates that, following the approach of Willenbrock and collaborators, the incorporation of novel genomes might produce a decay in the number of genes that compose the core genome of *S. maltophilia* (Willenbrock et al., 2007). **(C)** Representation of the pangenome obtained by analyzing 24 genomes of *S. maltophilia* isolates. Each circle represents the contribution of each genome to the composition analysis. Genes shared by several strains are clustered at the right side of the circle and strain-specific genes are clustered at its left side. The list of strains displays their names from the inner to the outer circle. \*Complete genomes used in this study. Red: clinical isolates. Green: environmental strains.

are analyzed (Figure 2C). It is important to notice that, since draft genomes are analyzed, the lack of genes in one specific strain may be the consequence of its presence at the edge of one contig, in which case will be annotated as a truncated gene,

although this putative truncation will be the consequence of the method of analysis, not of a real absence. Consequently, the "soft-core genome" (Kaas et al., 2012) was also analyzed. By using this approach, we estimated the number of orthologous genes





**FIGURE 3 |** Estimation of *S. maltophilia* core, soft-core, shell and cloud genome. **(A)** Distribution of the predicted CDSs along the 24 *S. maltophilia* genomes representing the core genome (blue bars), the soft-core genome (red bars), the shell genome (yellow bars), and the cloud genome (orange bars) of each of the strains is shown. **(B)** Representation of the overall distribution of genes contributing to the core, the soft, the shell and the cloud genomes. As shown, most genes not belonging to the core/soft-core genome are present in just one or two strains, suggesting that most of the *S. maltophilia* pangenome is strain specific, and does not depend on the environmental or clinical origin of the isolate.

shared by ~90% of the organisms included in the comparative analysis. Applying the soft-core genome concept, the number of orthologous clusters increased to 3045. When the 24 genomes were analyzed independently, we estimated that the size of the core genome for each *S. maltophilia* isolate comprised around 59.11% (minimum 54.6%; maximum 64%) of the CDS from each genome. Further, the analysis of the pangenome shows that most of the genes carried by *S. maltophilia* and not belonging to the core genome are strain-specific, suggesting specific adaptations for each isolate more than a common pattern of speciation of some members of the population toward virulence. Indeed, among those genes not belonging to the soft-core genome, and shared by 3–21 strains (dubbed as the ‘shell genome’), just 1337 gene clusters, from the total of 7108 orthologous genes present in the pangenome, were found, indicating that the vast majority of *S. maltophilia* genes, not belonging to its core genome, are strain-specific (Figure 3). The speciation of bacterial pathogens usually involve the acquisition by horizontal gene transfer (HGT) of virulence genes, followed by the loss of other genes and the selection of mutations that allow the fine tuning of the metabolism (Martinez, 2013), a process very well studied in the case of *Yersinia* (Achtman et al., 1999; Wren, 2003; Achtman et al., 2004; Zhou and Yang, 2009). HGT is the consequence of either transformation or either the acquisition of mobile elements. Once these mobile elements are acquired, they can be fixed or spread to other hosts, a situation highly relevant in the case of antibiotic resistance (Martinez et al., 2009, 2017). Despite that the presence of several genes in the cloud genome of *S. maltophilia* suggests that this process has largely contributed

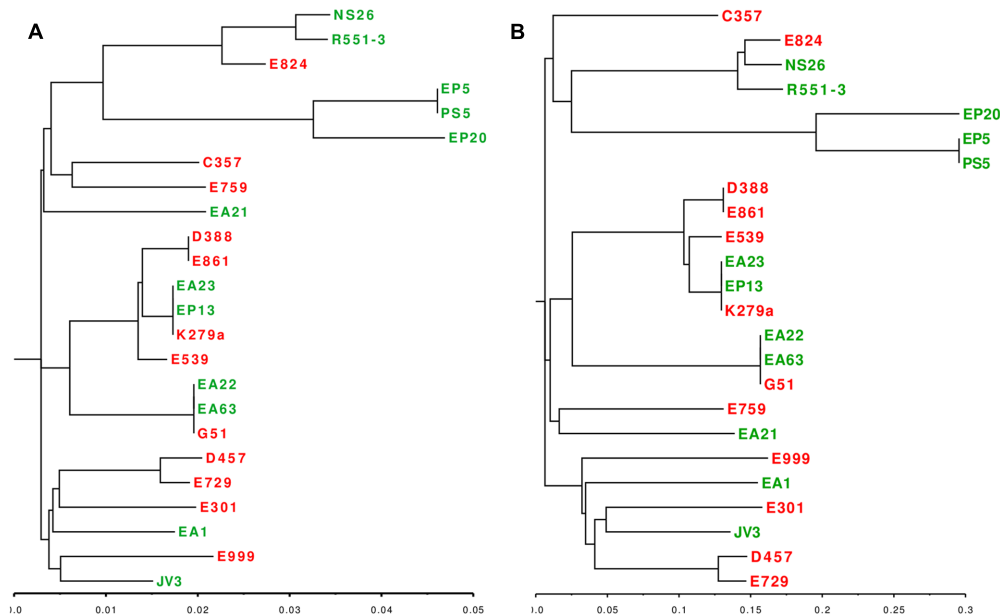
to the diversification of this pathogen, clear information on its mobilome has not been published. Indeed, only three whole sequenced *S. maltophilia* plasmids are available, which makes difficult to estimate the role of these mobile elements in the evolution of *S. maltophilia*.

In addition to the finding that there are several strain-specific genes, is important to recall that, as Figure 2C shows, a differential distribution of genes, not belonging to the core genome, was not found when environmental and clinical isolates of *S. maltophilia* were compared. This result further suggests that there is not a specific phylogenetic branch, deriving from the acquisition of a specific set of virulence determinants by the clinical *S. maltophilia* isolates, which can drive the speciation of this microorganism toward virulence.

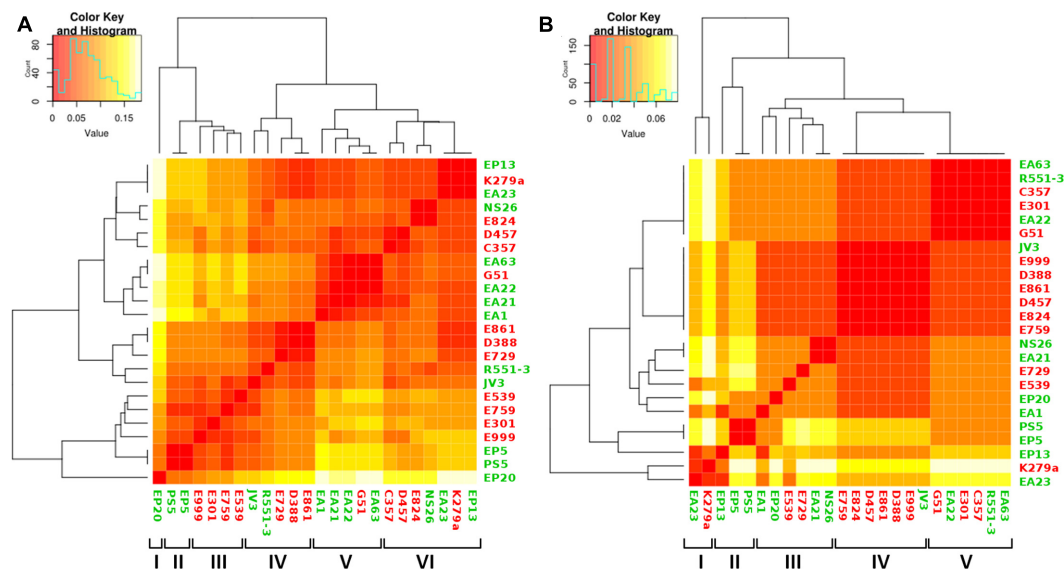
It is important to highlight that several articles analyzing bacterial core genomes make use of draft genomes in which genes at the edges of contigs are interrupted, which introduce some noise in the analysis that can produce an underestimation of the size of core genomes. Hence to avoid such noise, and since generation of complete genomes is by far more expensive than draft genomes, we propose using the soft-core genome as the right estimator of the number of genes that are common to all members of a given bacterial species.

### ***In Silico* Multi-Locus Sequences Typing (MSLT) and Core Genome SNPs**

Phylogenetic branches do not depend just on the presence/absence of genes, but in the fixation of specific mutations that can also provide differentiation of clinical and



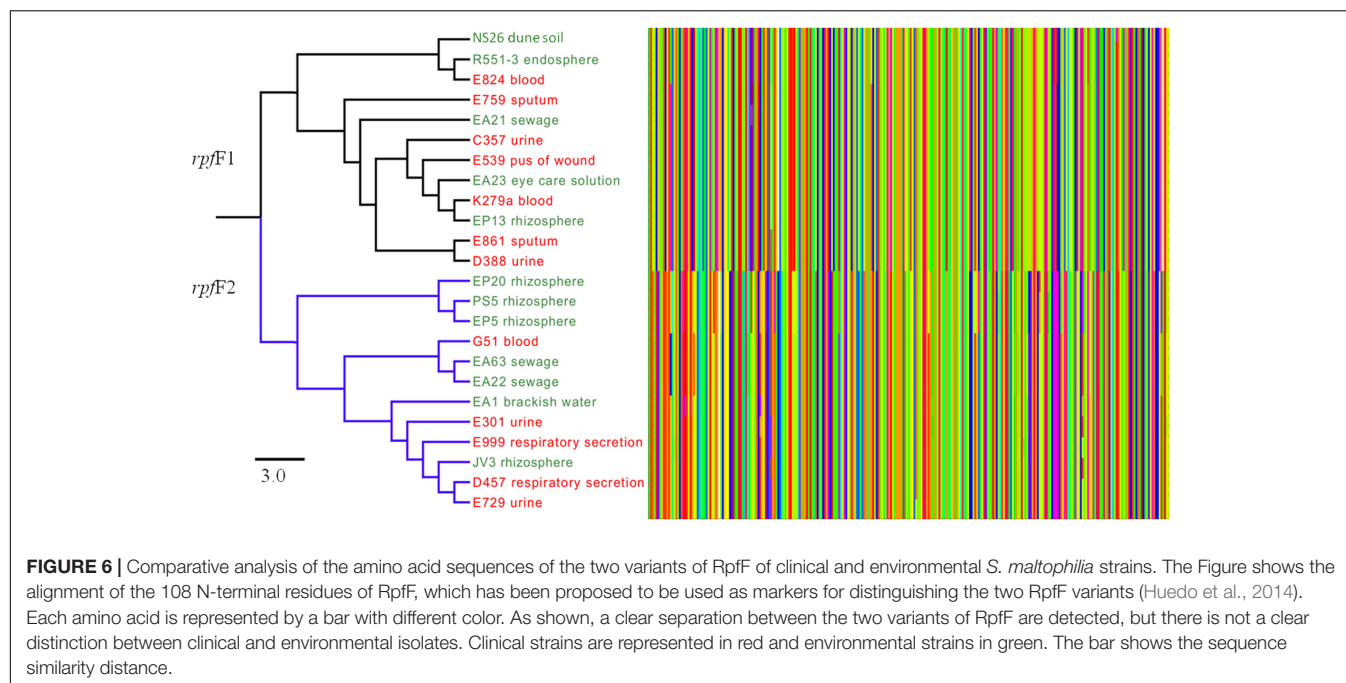
**FIGURE 4 |** Phylogenetic distribution of clinical and environmental *S. maltophilia* isolates. **(A)** Phylogenetic dendrogram based on the *in vitro* MLST analysis of seven concatenated genes (*atpD*, *guaA*, *nuoD*, *recA*, *gapA*, *mutM*, *ppsA*) **(B)** Phylogenetic dendrogram based on the alignment of SNPs present in the core genome of *S. maltophilia*. Red: clinical isolates. Green: environmental strains. As shown, both analysis grouped the strains in three major clusters, each one containing clinical and environmental strains.



**FIGURE 5 |** Analysis of genome composition of clinical and environmental *S. maltophilia* isolates based on functional categories. Color key represents the scale of similarities from red (high) to yellow (low), and counts axis is the number of observed pairs (x, y) that fall into each binary event (presence/absence of shared functional categories for each isolate) represented by the histograms (blue lines). Green, environmental isolates; red clinical isolates. **(A)** Heatmap showing the clustering of clinical and environmental *S. maltophilia* isolates based in the genes with functional roles classified at FIGfam within the subsystem 'Virulence, Disease and Defense' in the RAST server. The clustering, based on a presence/absence matrix, revealed that most clusters contain both clinical and environmental strains. **(B)** Heatmap showing the clustering of clinical and environmental *S. maltophilia* strains based in the presence/absence of a specific set of virulence determinants described in Adamek et al. (2014). As shown, most branches contain both environmental and clinical isolates.

environmental isolates in different phylogenetic branches. To address this possibility, we performed two types of complementary analysis, namely *in silico* MLST and study

of the core genome SNPs. Seven genes were used as markers for the MLST analysis: *atpD*, *gapA*, *guaA*, *mutM*, *nuoD*, *ppsA* and *recA*. The phylogenetic tree based on the alignment of



these genes consisted of three major groups, each one of them comprised by clinical and environmental strains (**Figure 4A**). All SNPs were identified using *S. maltophilia* K279a as reference and phylogenetic dendrogram based on the core SPNs alignment was consistent with the topology and branches of the MSLT-based tree (**Figure 4B**). The data combining the genotypic profiling provided by the MSLT and the evaluation of the core SNPs of the 24 strains presented in this study revealed that *S. maltophilia* is a diverse complex, forming an interlaced taxon, sharing the same attributes between clinical and environmental strains without preference with respect to their origin.

## Functional-Based Comparison between Clinical and Environmental Strains of *S. maltophilia*

Even though we did not find a clear distinction between the genome sizes and their CDS composition of clinical and environmental *S. maltophilia* strains, it is still possible that some functional categories, particularly those dealing with virulence are enriched as a function of the habitat (clinical or environmental) from which these strains have been isolated. Consequently, the 20 sequenced genomes and the four complete genomes used as reference were analyzed according to the functional groups of the CDS present in each of the genomes to further explore the relationship between habitat and genome composition. The presence of genes classified into the FIGfam subsystem 'Virulence, Disease and Defense' (Overbeek et al., 2005) was analyzed in all strains (Supplementary Table S1). From this information, a customized set of genes, containing only genes that were not present in all the isolates (Supplementary Table S2) was used to create a presence/absence matrix with roles not shared by all strains. Calculation of the average distance

of strains and further clustering indicated the formation of six hierarchical clusters (**Figure 5A**). From the six clusters, only clusters II and III were formed exclusively by strains isolated from the same habitat. Environmental strains isolated from the rhizosphere, EP5 and PS5, composed the cluster II and both lacked some functional roles attributed to copper resistance. Cluster III was constituted by the clinical isolates E999, E301, E759 and E539 isolated from respiratory secretion, urine, sputum and pus, respectively. These isolates did not present five functional roles responsible for copper resistance. Other three clusters presented both clinical and environmental isolates in their composition, and EP20 did not group with other strains. Cluster IV, despite its composition including clinical and environmental isolates, presented a subdivision in those branches, creating two distinct sub-clusters composed by two environmental strains isolated from plants, JV3 and R551-3. Clinical isolates E861 and E729 were obtained from patients presenting urinary infection and strain D388 obtained from blood sample. The isolates grouped in cluster V present different origins and were characterized by the presence of genes related with Hg resistance (Supplementary Table S2). The presence of these genes suggest that these strains are able to inactivate Hg toxic forms into less toxic compounds. Finally, cluster VI was characterized by the diversity of organisms isolates from different habitats.

Taking into consideration that most clusters contain clinical and environmental isolates, and that the observed differences do not involve the presence of specific virulence genes in the clinical isolates, our results reinforce the notion that there are not clear distinctions between clinical and environmental *S. maltophilia* strains, even when the analysis is based in the distribution of functional categories.

It is worth mentioning, however, that virulence can be due to the presence of a small subset of genes and global analysis would not be sufficient to distinguish the presence or absence of such genes. Consequently, we screened for the presence of a set of genes that has been described as markers for *S. maltophilia* virulence (Adamek et al., 2014) (Supplementary Table S3). By using this dataset, the presence of five clusters was shown with four of them mixing clinical and environmental strains from different origins (Figure 5B). Only cluster II contained exclusively environmental strains (three), all obtained from plant rhizosphere. These results reinforced the idea that the genomic composition is not sufficient to establish a clear separation between clinical and environmental strains of *S. maltophilia*. Cluster I grouped the isolates EA23 and K279a that presents genes encoding filamentous hemagglutinins, which are important for adhesion and spread of bacteria through the respiratory tract (Colombi et al., 2006; Crossman et al., 2008). Despite it was not grouped in the same cluster, the isolate EP13 presented as well filamentous hemagglutinins genes. Isolates EP5 and PS5 were clustered in the same branch, in agreement with their complete CDS composition. Seven isolates did not present five functional roles responsible for copper resistance: four of them, E759, E999, E301 and E539, were clinical strains that shared the same cluster when the analysis was performed using the classification based in functional roles (Figure 5A). Otherwise, they did not share the same cluster when analyzed using the

set of virulence factors (Figure 5B). The same happened with the environmental isolates PS5 and EP5 that shared the same clusters in both types of analysis. Altogether, the phylogenetical relationship of all strains analyzed in this study, calculated in base of their CDS composition and the clustering in orthologous groups, demonstrated that clinical and environmental strains did not form two independent evolutionary lineages. These results support the idea that clinical and environmental isolates are closely related and the pathogenic behavior does not depend on the acquisition of a specific set of virulence genes.

## Quorum-Sensing Signals

The quorum-sensing system (QS) is responsible for the synchronization of particular bacterial behaviors on a population scale. In the case of *S. maltophilia* this process is relevant for *S. maltophilia* virulence and for its interaction with plants (Alavi et al., 2013, 2014), and depends on the Diffusible Signal Factor QS (DSF-QS), which has been identified as the fatty acid *cis*-11-methyl-2-dodecenoic acid (Fouhy et al., 2007).

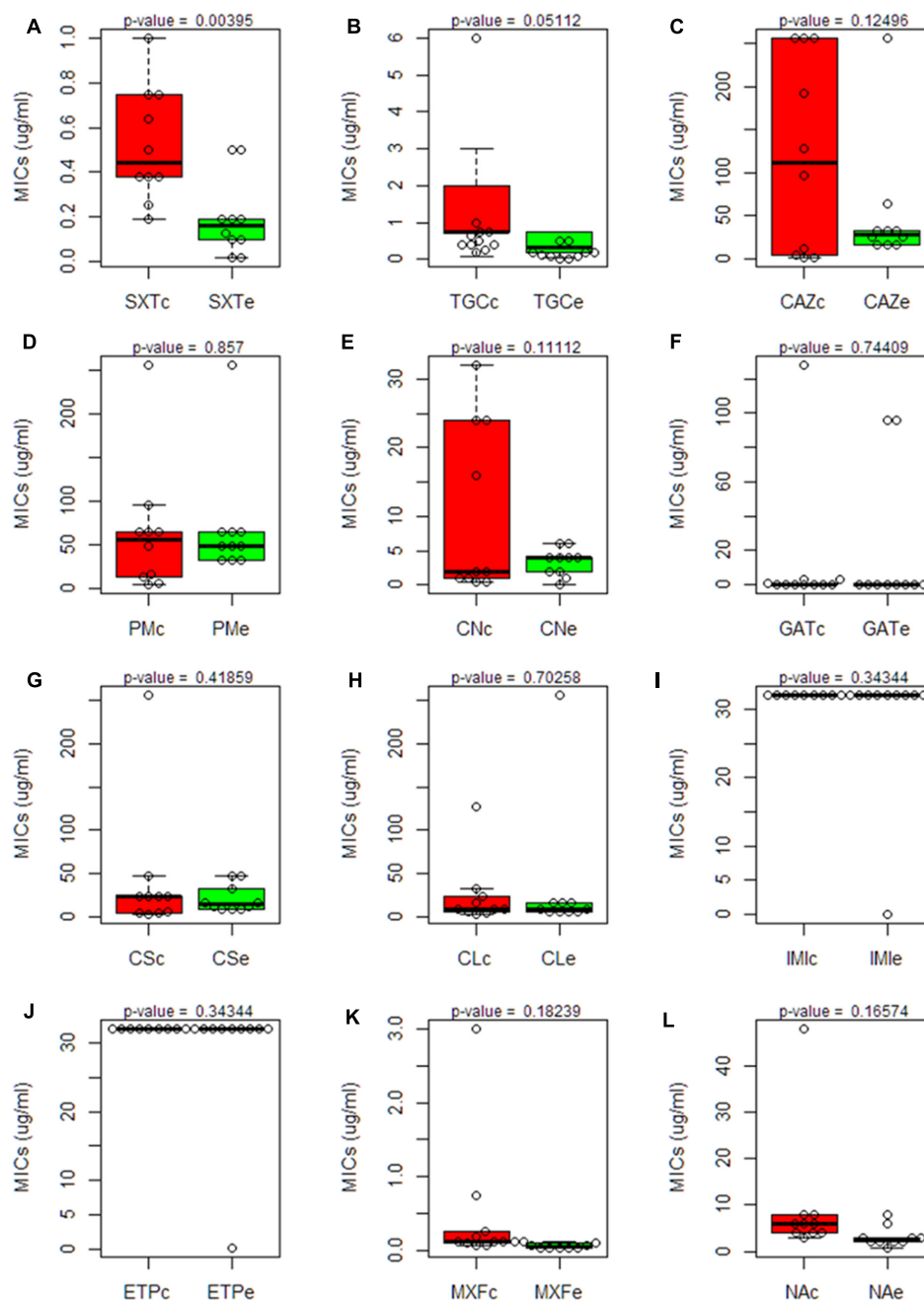
In a previous study the existence of two different alleles for the *rpff* gene, that is essential for the synthesis of DSF has been described. Each of the alleles defined a branch presenting a different virulence behavior (Huedo et al., 2014). It is then still possible that environmental and clinical isolates could present a differential virulence based in the presence/absence of a specific *rpff* allele. Since these variants are markers of two different

**TABLE 3 |** Minimal Inhibitory Concentrations (MICs) of 20 new sequenced strains and of the model strain *S. maltophilia* D457.

	Strains	SXT	TGC	CAZ*	PM*	CN	GAT	CS	CL*	IMI*	ETP*	MXF	NA
Clinical strains	E729	0.75	3	1	4	2	1	4	24	>32	>32	0.75	8
	E759	1	2	>256	>256	1	128	6	32	>32	>32	0.25	4
	E999	0.38	2	4	12	24	0.25	4	16	>32	>32	0.13	8
	G51	0.38	0.75	>256	48	2	0.06	24	3	>32	>32	0.09	6
	E301	0.19	0.75	1.5	6	0.38	0.13	24	4	>32	>32	0.19	6
	D388	0.25	0.75	96	64	24	0.13	24	6	>32	>32	0.06	3
	E861	0.38	0.5	128	64	16	0.13	24	8	>32	>32	0.13	4
	C357	0.75	6	192	96	32	3	48	128	>32	>32	3	48
	E539	0.5	0.75	256	64	1	0.13	256	8	>32	>32	0.13	6
	E824	0.64	0.09	12	16	0.5	0.05	3	8	>32	>32	0.06	4
Environmental strains	NS26	0.19	0.75	16	64	2	0.13	12	6	>32	>32	0.13	3
	EP13	0.09	0.19	16	32	4	96	12	16	>32	>32	0.13	0.8
	EA22	0.5	0.75	32	48	2	0.13	32	6	>32	>32	0.03	3
	EA1	0.5	0.19	32	48	4	0.02	8	16	>32	>32	0.03	2
	PS5	0.02	0.05	64	>256	0.09	0.02	16	>256	0.06	0.2	0.03	8
	EA23	0.19	0.25	24	64	4	0.19	8	6	>32	>32	0.09	2
	EP20	0.09	0.19	16	32	6	0.09	48	8	>32	>32	0.03	2
	EP5	0.02	0.75	24	32	4	96	8	6	>32	>32	0.03	2
	EA21	0.19	0.5	>256	64	1	0.09	16	16	>32	>32	0.06	3
	EA63	0.13	0.38	32	48	6	0.06	48	8	>32	>32	0.06	6
	D457	0.13	1.5	1.5	16	6	0.5	32	12	32	32	0.25	8
	MIC50	0.25	0.75	32	48	2	0.13	16	8	32	32	0.09	4
	MIC90	0.75	2	256	96	24	96	48	32	32	32	0.25	8

Values of the MIC50 (Minimum Inhibitory Concentration required to inhibit the growth of 50% of the isolates) and MIC90 (Minimum Inhibitory Concentration required to inhibit the growth of 90% of the isolates). Antibiotics: SXT, Trimethoprim/Sulfamethoxazole; TGC, Tigecyclin; CAZ, Ceftazidime; PM, Cefepime; CN, Gentamicin; GAT, Gatifloxacin; CS, Colistin; CL, Chloramphenicol; IM, Imipenem; ETP, Ertapenem; MXF, Moxifloxacin; NA, Nalidixic Acid. \*At least one strain presented a MIC above the highest antibiotic concentration in the strip test.

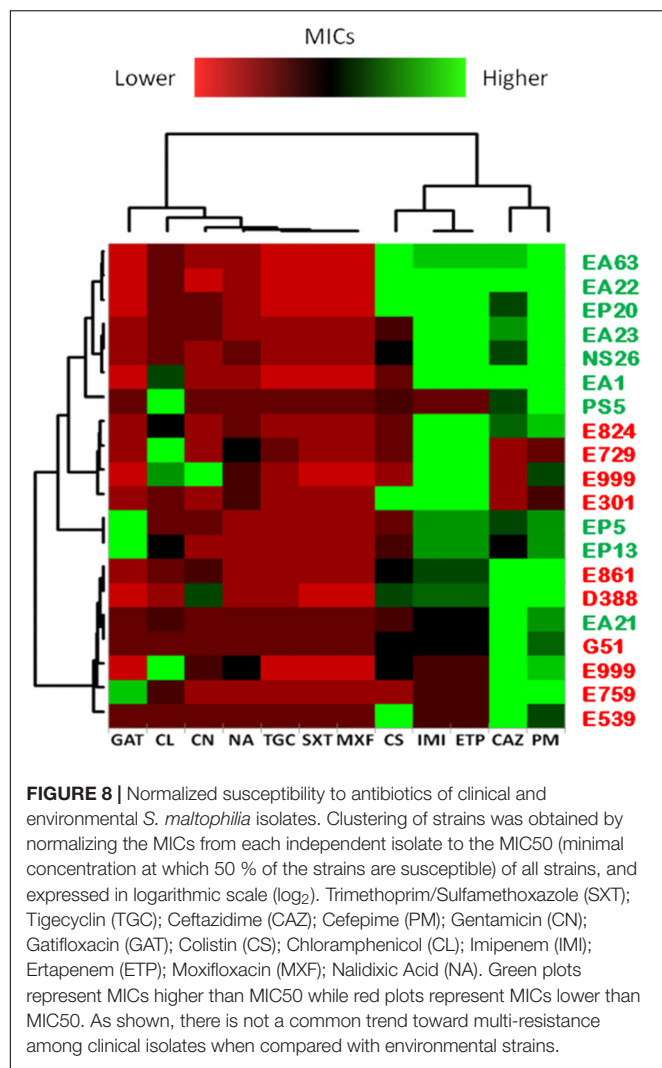




**FIGURE 7 |** Comparison of the susceptibility to antibiotics of clinical and environmental *S. maltophilia* isolates. Boxplot charts representing the Minimal Inhibitory Concentrations (MICs) for all clinical and environmental isolates obtained using antibiogram strip-tests of 12 antibiotics from different families: (SXT, Trimethoprim/Sulfamethoxazole; TGC, Tigecyclin; CAZ, Ceftazidime; PM, Cefepime; CN, Gentamicin; GAT, Gatifloxacin; CS, Colistin; CL, Chloramphenicol; IMI, Imipenem; ETP, Ertapenem; MXF, Moxifloxacin; NA, Nalidixic Acid). The median and the quartiles for the MIC values in each group are shown. Clinical isolates are represented in red box plots; environmental isolates are represented in green box plots. Statistical significance of the results was estimated by using the *t*-Student test. A significant difference ( $p$ -value  $< 0.05$ ) was found just in the case of Trimethoprim/Sulfamethoxazole. Trimethoprim/Sulfamethoxazole (SXT); Tigecyclin (TGC); Ceftazidime (CAZ); Cefepime (PM); Gentamicin (CN); Gatifloxacin (GAT); Colistin (CS); Chloramphenicol (CL); Imipenem (IMI); Ertapenem (ETP); Moxifloxacin (MXF); Nalidixic Acid (NA). Each panel, from (A–L) represents the MICs of one antibiotic.

phylogenetic branches, each one presenting differences in terms of virulence, we analyzed their presence in the 24 studied genomes. Using the available sequences for *rpfF1* and *rpfF2*, a

direct search was performed for the corresponding DNA region of *rpfF* in all 24 genomes. All 24 strains of *S. maltophilia* harbor this gene with different length and variable residues along the



sequence. In agreement with previous results (Huedo et al., 2014), the studied isolates are distributed into two distinct groups, each one presenting a different *rpfF* variant. Each group comprised 12 strains (Figure 6); however, there was not a clear difference in the distribution of clinical and environmental isolates between both groups. The cluster with the *RpfF1* variant, which displays detectable DSF production (Huedo et al., 2014), comprised four environmental strains and eight clinical isolates, while the cluster containing the *RpfF2* allele, with no significant effect on virulence-related phenotypes, presented seven environmental and five clinical strains.

## Antibiotic Susceptibility of Clinical and Environmental Isolates of *S. maltophilia*

The Minimal Inhibitory Concentrations (MIC) of 12 different antibiotics, belonging to a wide range of structural families and presenting different targets, were established for the 20 isolates. The strain D457, which has been used in several studies on antibiotic resistance in *S. maltophilia* (Alonso and Martinez, 1997) was included as a control. The results (Table 3) were plotted

in quartiles. As shown in Figure 7, the clinical isolates, as a group, present a trend toward higher levels of resistance than the environmental ones.

The environmental strain PS5 was the only isolate susceptible to IMI and ETP, while the other strains grew over the maximum value of this strip-test ( $>32 \mu\text{g/ml}$ ), a feature that fits with previous information showing that *S. maltophilia* is resistant to these antibiotics (Howe et al., 1997). Notably, the same isolate, PS5, presented the highest level of resistance to CL ( $>256 \mu\text{g/ml}$ ), followed by the strain C357 (MIC  $128 \mu\text{g/ml}$ ). For the other strains, the values ranged between 3 and  $32 \mu\text{g/ml}$ . This may suggest that all *S. maltophilia* isolates, independently from their origin, present similar chances to acquire resistance to this antibiotic.

Although all isolates displayed low susceptibility to the tested antibiotics, when we analyze just the antibiotic concentration ranges where the values did not exceed the maximum concentration of the strip tests, the clinical strains presented overall higher MIC values for the antibiotics SXT, TGC, GAT, MXF and NA when compared with the environmental isolates (Figure 7). Nevertheless, this difference was statistically significant only in the case of STX. Therefore, despite there seems to be a trend toward lower MIC values in the environmental isolates, and in agreement with other studies (Berg et al., 1999), the multiple-antibiotic-resistance pattern of both clinical and environmental strains does not present significant differences and might be explained by the intrinsic resistome linked to the core genome of this species.

The previous analysis was based in the independent analysis of each of the antibiotics in the full population. To analyze a different aspect of the problem: the susceptibility to several antibiotics in each independent isolate, further comparisons of the clinical and environmental isolates were performed normalizing the obtained MICs by the MIC50 of all strains (Figure 8). Normalization of the MICs by the MIC50 of the 20 isolates grouped the environmental strains NS26, EA22, EA1, PS5, EA23, EP20 and EA63 in one branch presenting overall less susceptibility to carbapenems, imipenem and ertapenem, and the cephalosporins, ceftazidime and cefepime, than the other strains, indicating that, at least for some antibiotics, environmental isolates can present higher levels of resistance than clinical strains. Previous publications have shown that both clinical and environmental *S. maltophilia* isolates are highly resistant to antibiotics (Berg et al., 1999). Our results confirm this issue: the MICs of most antibiotics are high in all isolates as compared with other bacterial pathogens, and in occasions environmental strains are even less susceptible than clinical isolates.

## CONCLUSION

When looking to the structure of bacterial species presenting infective and non-infective habitats, three situations can be foreseen. Either the species present specific virulence branches, as it happens in the case of *Escherichia coli*, either all isolates can produce an infection in healthy and sick people as in the case of *Yersinia pestis*, either all isolates can produce infection,

but only in people with a previous basal disease, as it has been described in the case of *P. aeruginosa*. The consequences in terms of preventing infections by each one of these species would be different. For the first type of microorganisms, surveillance must be taken at the clonal level: some clones constitute a risk while some others are not dangerous. For the second, each member of the species must be considered as a risk for human health. In the third case, the risk is not mainly due to the organism itself, which does not infect the community, but to the situation of the potential host to be infected. Our results indicate that *S. maltophilia* belongs to the third category; all strains are likely equivalent in their capability of infecting humans, but only patients presenting severe underlying diseases including cystic fibrosis would be infected by this pathogen. Given the high biotechnological potential of *S. maltophilia*, both for confined and non-confined applications, there are concerns on the risk that this use may have for human health. Our results indicate that this concern applies just for people with underlying diseases and not for the community and, given that *S. maltophilia* is an environmental ubiquitous and cosmopolitan organism, its use in the habitats that this bacterium regularly colonizes will likely produce just an incremental risk of acquiring infections, even in the case of patients presenting underlying diseases.

A final reflection concerns the distribution of *S. maltophilia* pangenome. Most genes not belonging to the core genome are present in just one or a few strains. Together with the finding that *S. maltophilia* presents an open genome, this suggests that

*S. maltophilia* can likely colonize a full range of microniches and, for such colonization, each member of this bacterial species is capable of acquiring a specific set of genes through HGT.

## AUTHOR CONTRIBUTIONS

FL performed the experiments and bioinformatic analysis presented in the work. All authors contributed to the design and interpretation of the results, as well as to writing the article, and approved it for publication.

## FUNDING

Work in our laboratory is supported by grants from the Instituto de Salud Carlos III (Spanish Network for Research on Infectious Diseases [RD16/0016/0011]) and from the Spanish Ministry of Economy and Competitiveness (BIO2014-54507-R and JPI Water StARE JPIW2013-089-C02-01). FL was the recipient of La Caixa fellowship.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02190/full#supplementary-material>

## REFERENCES

- Achtman, M., Morelli, G., Zhu, P., Wirth, T., Diehl, I., Kusecek, B., et al. (2004). Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17837–17842. doi: 10.1073/pnas.0408026101
- Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiry, A., and Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14043–14048. doi: 10.1073/pnas.96.24.14043
- Adamek, M., Linke, B., and Schwartz, T. (2014). Virulence genes in clinical and environmental *Stenotrophomonas maltophilia* isolates: a genome sequencing and gene expression approach. *Microb. Pathog.* 67–68, 20–30. doi: 10.1016/j.micpath.2014.02.001
- Alavi, P., Muller, H., Cardinale, M., Zachow, C., Sanchez, M. B., Martinez, J. L., et al. (2013). The DSF quorum sensing system controls the positive influence of *Stenotrophomonas maltophilia* on plants. *PLOS ONE* 8:e67103. doi: 10.1371/journal.pone.0067103
- Alavi, P., Starcher, M. R., Thallinger, G. G., Zachow, C., Muller, H., and Berg, G. (2014). *Stenotrophomonas* comparative genomics reveals genes and functions that differentiate beneficial and pathogenic bacteria. *BMC Genomics* 15:482. doi: 10.1186/1471-2164-15-482
- Alonso, A., and Martinez, J. L. (1997). Multiple antibiotic resistance in *Stenotrophomonas maltophilia*. *Antimicrob. Agents Chemother.* 41, 1140–1142.
- Alonso, A., and Martinez, J. L. (2001). Expression of multidrug efflux pump SmeDEF by clinical isolates of *Stenotrophomonas maltophilia*. *Antimicrob. Agents Chemother.* 45, 1879–1881. doi: 10.1128/AAC.45.6.1879-1881.2001
- Alonso, A., Rojo, F., and Martinez, J. L. (1999). Environmental and clinical isolates of *Pseudomonas aeruginosa* show pathogenic and biodegradative properties irrespective of their origin. *Environ. Microbiol.* 1, 421–430. doi: 10.1046/j.1462-2920.1999.00052.x
- Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Berg, G., Marten, P., and Ballin, G. (1996). *Stenotrophomonas maltophilia* in the rhizosphere of oilseed rape occurrence, characterization and interaction with phytopathogenic fungi. *Microbiol. Res.* 151, 18–27. doi: 10.1016/S0944-5013(96)80051-6
- Berg, G., and Martinez, J. L. (2015). Friends or foes: can we make a distinction between beneficial and harmful strains of the *Stenotrophomonas maltophilia* complex? *Front. Microbiol.* 6:241. doi: 10.3389/fmicb.2015.00241
- Berg, G., Roskot, N., and Smalla, K. (1999). Genotypic and phenotypic relationships between clinical and environmental isolates of *Stenotrophomonas maltophilia*. *J. Clin. Microbiol.* 37, 3594–3600.
- Brooke, J. S. (2012). *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin. Microbiol. Rev.* 25, 2–41. doi: 10.1128/CMR.00019-11
- Brooke, J. S., Di Bonaventura, G., Berg, G., and Martinez, J. L. (2017). Editorial: a multidisciplinary look at *Stenotrophomonas maltophilia*: an emerging multi-drug-resistant global opportunistic pathogen. *Front. Microbiol.* 8:1511. doi: 10.3389/fmicb.2017.01511
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Chang, Y. T., Lin, C. Y., Chen, Y. H., and Hsueh, P. R. (2015). Update on infections caused by *Stenotrophomonas maltophilia* with particular attention to resistance mechanisms and therapeutic options. *Front. Microbiol.* 6:893. doi: 10.3389/fmicb.2015.00893

- Chiarini, L., Bevivino, A., Dalmastri, C., Tabacchioni, S., and Visca, P. (2006). *Burkholderia cepacia* complex species: health hazards and biotechnological potential. *Trends Microbiol.* 14, 277–286. doi: 10.1016/j.tim.2006.04.006
- Colombi, D., Oliveira, M. L., Campos, I. B., Monedero, V., Perez-Martinez, G., and Ho, P. L. (2006). Haemagglutination induced by *Bordetella pertussis* filamentous haemagglutinin adhesin (FHA) is inhibited by antibodies produced against FHA(430-873) fragment expressed in *Lactobacillus casei*. *Curr. Microbiol.* 53, 462–466. doi: 10.1007/s00284-005-0388-0
- Contreras-Moreira, B., and Vinuesa, P. (2013). GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13
- Criscuolo, A., and Brisse, S. (2013). AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* 102, 500–506. doi: 10.1016/j.ygeno.2013.07.011
- Crossman, L. C., Gould, V. C., Dow, J. M., Vernikos, G. S., Okazaki, A., Sebaihia, M., et al. (2008). The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.* 9:R74. doi: 10.1186/gb-2008-9-4-r74
- Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- De Boer, W., Klein Gunnewiek, P. J., Kowalchuk, G. A., and Van Veen, J. A. (2001). Growth of chitinolytic dune soil beta-subclass *Proteobacteria* in response to invading fungal hyphae. *Appl. Environ. Microbiol.* 67, 3358–3362. doi: 10.1128/AEM.67.8.3358-3362.2001
- Dungan, R. S., Yates, S. R., and Frankenberger, W. T. Jr. (2003). Transformations of selenate and selenite by *Stenotrophomonas maltophilia* isolated from a seleniferous agricultural drainage pond sediment. *Environ. Microbiol.* 5, 287–295. doi: 10.1046/j.1462-2920.2003.00410.x
- Dunne, C., Moenne-Loccoz, Y., De Bruijn, F. J., and O'gara, F. (2000). Overproduction of an inducible extracellular serine protease improves biological control of *Pythium ultimum* by *Stenotrophomonas maltophilia* strain W81. *Microbiology* 146(Pt 8), 2069–2078. doi: 10.1099/00221287-146-8-2069
- Esposito, A., Pompilio, A., Bettua, C., Crocetta, V., Giacobazzi, E., Fiscarelli, E., et al. (2017). Evolution of *Stenotrophomonas maltophilia* in cystic fibrosis lung over chronic infection: a genomic and phenotypic population study. *Front. Microbiol.* 8:1590. doi: 10.3389/fmicb.2017.01590
- Fouhy, Y., Scanlon, K., Schouest, K., Spillane, C., Crossman, L., Avison, M. B., et al. (2007). Diffusible signal factor-dependent cell-cell signaling and virulence in the nosocomial pathogen *Stenotrophomonas maltophilia*. *J. Bacteriol.* 189, 4964–4968. doi: 10.1128/JB.00310-07
- Gherardi, G., Creti, R., Pompilio, A., and Di Bonaventura, G. (2015). An overview of various typing methods for clinical epidemiology of the emerging pathogen *Stenotrophomonas maltophilia*. *Diagn. Microbiol. Infect. Dis.* 81, 219–226. doi: 10.1016/j.diagmicrobio.2014.11.005
- Howe, R. A., Wilson, M. P., Walsh, T. R., and Millar, M. R. (1997). Susceptibility testing of *Stenotrophomonas maltophilia* to carbapenems. *J. Antimicrob. Chemother.* 40, 13–17. doi: 10.1093/jac/40.1.13
- Huedo, P., Yero, D., Martinez-Servat, S., Estibariz, I., Planell, R., Martinez, P., et al. (2014). Two different rpf clusters distributed among a population of *Stenotrophomonas maltophilia* clinical strains display differential diffusible signal factor production and virulence regulation. *J. Bacteriol.* 196, 2431–2442. doi: 10.1128/JB.001540-14
- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Jakobi, M., Winkelmann, G., Kaiser, D., Kempler, C., Jung, G., Berg, G., et al. (1996). Maltophilin: a new antifungal compound produced by *Stenotrophomonas maltophilia* R3089. *J. Antibiot.* 49, 1101–1104. doi: 10.7164/antibiotics.49.1101
- Jeon, Y. D., Jeong, W. Y., Kim, M. H., Jung, I. Y., Ahn, M. Y., Ann, H. W., et al. (2016). Risk factors for mortality in patients with *Stenotrophomonas maltophilia* bacteremia. *Medicine* 95:e4375. doi: 10.1097/MD.0000000000004375
- Kaas, R. S., Friis, C., Ussery, D. W., and Aarestrup, F. M. (2012). Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* 13:577. doi: 10.1186/1471-2164-13-577
- Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719. doi: 10.1093/nar/gkn668
- Langfelder, P., and Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* 46, 1–17. doi: 10.18637/jss.v046.i11
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11
- Lee, E. Y., Jun, Y. S., Cho, K. S., and Ryu, H. W. (2002). Degradation characteristics of toluene, benzene, ethylbenzene, and xylene by *Stenotrophomonas maltophilia* T3-c. *J. Air Waste Manag. Assoc.* 52, 400–406. doi: 10.1080/10473289.2002.10470796
- Lira, F., Hernandez, A., Belda, E., Sanchez, M. B., Moya, A., Silva, F. J., et al. (2012). Whole-genome sequence of *Stenotrophomonas maltophilia* D457, a clinical isolate and a model strain. *J. Bacteriol.* 194, 3563–3564. doi: 10.1128/JB.00602-12
- Lucas, S., Copeland, A., Lapidus, A., Glavina Del Rio, T., Dalin, E., and Tice, H. (2008). Complete sequence of *Stenotrophomonas maltophilia* R551-3. *EMBL/GenBank/DBJ Databases*.
- Lucas, S., Han, J., Lapidus, A., Cheng, J.-F., Goodwin, L., Pitluck, S., et al. (2011). Complete sequence of *Stenotrophomonas maltophilia* JV3. *EMBL/GenBank/DBJ Databases*.
- Martinez, J. L. (2013). Bacterial pathogens: from natural ecosystems to human hosts. *Environ. Microbiol.* 15, 325–333. doi: 10.1111/j.1462-2920.2012.02837.x
- Martinez, J. L., Coque, T. M., Lanza, V. F., De La Cruz, F., and Baquero, F. (2017). Genomic and metagenomic technologies to explore the antibiotic resistance mobilome. *Ann. N. Y. Acad. Sci.* 1388, 26–41. doi: 10.1111/nyas.13282
- Martinez, J. L., Fajardo, A., Garmendia, L., Hernandez, A., Linares, J. F., Martinez-Solano, L., et al. (2009). A global view of antibiotic resistance. *FEMS Microbiol. Rev.* 33, 44–65. doi: 10.1111/j.1574-6976.2008.00142.x
- Minkwitz, A., and Berg, G. (2001). Comparison of antifungal activities and 16S ribosomal DNA sequences of clinical and environmental isolates of *Stenotrophomonas maltophilia*. *J. Clin. Microbiol.* 39, 139–145. doi: 10.1128/JCM.39.1.139-145.2001
- Morales, G., Wiehlmann, L., Gudowius, P., Van Delden, C., Tummeler, B., Martinez, J. L., et al. (2004). Structure of *Pseudomonas aeruginosa* populations analyzed by single nucleotide polymorphism and pulsed-field gel electrophoresis genotyping. *J. Bacteriol.* 186, 4228–4237. doi: 10.1128/JB.186.13.4228-4237.2004
- Mukherjee, P., and Roy, P. (2016). Genomic potential of *Stenotrophomonas maltophilia* in bioremediation with an assessment of its multifaceted role in our environment. *Front. Microbiol.* 7:967. doi: 10.3389/fmicb.2016.00967
- Nangia, Y., Wangoo, N., Goyal, N., Shekhawat, G., and Suri, C. R. (2009). A novel bacterial isolate *Stenotrophomonas maltophilia* as living factory for synthesis of gold nanoparticles. *Microb. Cell Fact.* 8:39. doi: 10.1186/1475-2859-8-39
- Nicoletti, M., Iacobino, A., Prosseda, G., Fiscarelli, E., Zarrilli, R., De Carolis, E., et al. (2011). *Stenotrophomonas maltophilia* strains from cystic fibrosis patients: genomic variability and molecular characterization of some virulence determinants. *Int. J. Med. Microbiol.* 301, 34–43. doi: 10.1016/j.ijmm.2010.07.003
- Ochoa-Sanchez, L. E., and Vinuesa, P. (2017). Evolutionary genetic analysis uncovers multiple species with distinct habitat preferences and antibiotic resistance phenotypes in the *Stenotrophomonas maltophilia* complex. *Front. Microbiol.* 8:1548. doi: 10.3389/fmicb.2017.01548
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866
- Pompilio, A., Crocetta, V., Ghosh, D., Chakrabarti, M., Gherardi, G., Vitali, L. A., et al. (2016). *Stenotrophomonas maltophilia* phenotypic and genotypic diversity during a 10-year colonization in the lungs of a cystic fibrosis patient. *Front. Microbiol.* 7:1551. doi: 10.3389/fmicb.2016.01551



- Ribbeck-Busch, K., Roder, A., Hasse, D., De Boer, W., Martinez, J. L., Hagemann, M., et al. (2005). A molecular biological protocol to distinguish potentially human pathogenic *Stenotrophomonas maltophilia* from plant-associated *Stenotrophomonas rhizophila*. *Environ. Microbiol.* 7, 1853–1858. doi: 10.1111/j.1462-2920.2005.00928.x
- Rocco, F., De Gregorio, E., Colonna, B., and Di Nocera, P. P. (2009). *Stenotrophomonas maltophilia* genomes: a start-up comparison. *Int. J. Med. Microbiol.* 299, 535–546. doi: 10.1016/j.ijmm.2009.05.004
- Sanchez, M. B. (2015). Antibiotic resistance in the opportunistic pathogen *Stenotrophomonas maltophilia*. *Front. Microbiol.* 6:658. doi: 10.3389/fmicb.2015.00658
- Sanchez, M. B., Hernandez, A., and Martinez, J. L. (2009). *Stenotrophomonas maltophilia* drug resistance. *Future Microbiol.* 4, 655–660. doi: 10.2217/fmb.09.45
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026
- Suckstorff, I., and Berg, G. (2003). Evidence for dose-dependent effects on plant growth by *Stenotrophomonas* strains from different origins. *J. Appl. Microbiol.* 95, 656–663. doi: 10.1046/j.1365-2672.2003.02021.x
- Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* 42, D553–D559. doi: 10.1093/nar/gkt1274
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wiehlmann, L., Wagner, G., Cramer, N., Siebert, B., Gudowius, P., Morales, G., et al. (2007). Population structure of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8101–8106. doi: 10.1073/pnas.0609213104
- Willenbrock, H., Hallin, P. F., Wassenaar, T. M., and Ussery, D. W. (2007). Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. *Genome Biol.* 8:R267. doi: 10.1186/gb-2007-8-12-r267
- Wren, B. W. (2003). The yersiniae—a model genus to study the rapid evolution of bacterial pathogens. *Nat. Rev. Microbiol.* 1, 55–64. doi: 10.1038/nrmicro730
- Youenou, B., Favre-Bonte, S., Bodilis, J., Brothier, E., Dubost, A., Muller, D., et al. (2015). Comparative genomics of environmental and clinical *Stenotrophomonas maltophilia* strains with different antibiotic resistance profiles. *Genome Biol. Evol.* 7, 2484–2505. doi: 10.1093/gbe/evv161
- Yu, D., Yin, Z., Li, B., Jin, Y., Ren, H., Zhou, J., et al. (2016). Gene flow, recombination, and positive selection in *Stenotrophomonas maltophilia*: mechanisms underlying the diversity of the widespread opportunistic pathogen. *Genome* 59, 1063–1075. doi: 10.1139/gen-2016-0073
- Zhou, D., and Yang, R. (2009). Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect. Immun.* 77, 2242–2250. doi: 10.1128/IAI.01477-08
- Zonaro, E., Lampis, S., Turner, R. J., Qazi, S. J., and Vallini, G. (2015). Biogenic selenium and tellurium nanoparticles synthesized by environmental microbial isolates efficaciously inhibit bacterial planktonic cultures and biofilms. *Front. Microbiol.* 6:584. doi: 10.3389/fmicb.2015.00584

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Lira, Berg and Martínez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Secondary Bacterial Infections Associated with Influenza Pandemics

Denise E. Morris<sup>1\*</sup>, David W. Cleary<sup>1</sup> and Stuart C. Clarke<sup>1,2,3\*</sup>

<sup>1</sup> Infectious Disease Epidemiology Group, Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine, Institute for Life Sciences, University of Southampton, University Hospital Southampton Foundation NHS Trust, Southampton, United Kingdom, <sup>2</sup> Global Health Research Institute, University of Southampton, Southampton, United Kingdom, <sup>3</sup> NIHR Southampton Respiratory Biomedical Research Unit, Southampton, United Kingdom

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Universiti Sultan Zainal Abidin,  
Malaysia

### Reviewed by:

Jeanette Teo,  
National University Hospital,  
Singapore  
Siomar De Castro Soares,  
Universidade Federal do Triângulo  
Mineiro, Brazil  
Chanwit Tribudharat,  
Siriraj Hospital, Thailand

### \*Correspondence:

Stuart C. Clarke  
s.c.clarke@soton.ac.uk  
Denise E. Morris  
d.e.morris@soton.ac.uk

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 04 April 2017

**Accepted:** 24 May 2017

**Published:** 23 June 2017

### Citation:

Morris DE, Cleary DW and  
Clarke SC (2017) Secondary Bacterial  
Infections Associated with Influenza  
Pandemics. *Front. Microbiol.* 8:1041.  
doi: 10.3389/fmicb.2017.01041

Lower and upper respiratory infections are the fourth highest cause of global mortality (Lozano et al., 2012). Epidemic and pandemic outbreaks of respiratory infection are a major medical concern, often causing considerable disease and a high death toll, typically over a relatively short period of time. Influenza is a major cause of epidemic and pandemic infection. Bacterial co/secondary infection further increases morbidity and mortality of influenza infection, with *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus* reported as the most common causes. With increased antibiotic resistance and vaccine evasion it is important to monitor the epidemiology of pathogens in circulation to inform clinical treatment and development, particularly in the setting of an influenza epidemic/pandemic.

**Keywords:** influenza, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Staphylococcus aureus*, pandemic

## INTRODUCTION

From the Plague of Athens to the present day, infectious disease has beset mankind throughout history. Medical and socio-economic advances have substantially reduced this burden, the eradication of smallpox in 1979 (World Health Organization, 2017) and the remarkable successes against polio and parasitic Guinea worm disease being three examples of an extensive list. Respiratory tract infections, however, continue to be a major cause of morbidity and mortality worldwide (Lozano et al., 2012; Morse et al., 2012; Zumla et al., 2014). When combined, lower and upper respiratory infections are the fourth highest cause of global mortality (Lozano et al., 2012). Epi- and pandemic outbreaks of respiratory infection are a major medical concern, often causing considerable disease and a high death toll, typically over a relatively short period of time. The unpredictable nature of these outbreaks, in terms of their etiology and the reservoirs from which they emerge, the constant emergence of new antigenic variants by mutation, combined with transmission within potentially immunologically naïve populations facilitates the characteristic high proficiency of spread (Morse et al., 2012).

It is well established that both animals and humans can act as reservoirs of infection within which pathogens may adapt and evolve. Examples include *Coxiella burnetii* which typically causes Q fever in cattle, sheep and goats but can also infect humans (Eldin et al., 2017), the plague causing *Yersinia pestis*, infamously transmitted to humans by rats via a flea vector (Yang et al., 2016b), human immunodeficiency virus (HIV) which originated in non-human primates before spreading into the human population (Rupp et al., 2016) and of course the most common example, influenza, which circulates within and between swine, avian and human hosts (amongst others). This cross-species flow can lead to adaptations that result in an increased pathogenicity to susceptible hosts, creating the potential for localized outbreaks or global spread (Murphy, 1998;

Karesh et al., 2012; Morse et al., 2012). Important evolutionary modifications can occur during the timespan of an individual infection, permitting new and evolved strains of pathogens to emerge at an increased rate (Karesh et al., 2012). The evolution of pathogens (particularly zoonotic pathogens which account for 60% of human infectious diseases), and development of pandemics and epidemics, can be described in ecological principles whereby changing environmental pressures or opportunities drive a pathogen to exploit new niches or hosts to survive and thrive. This evolution is influenced by a range of anthropogenic factors, which include population expansion, changing land use and habitat destruction, selective pressures of increased antimicrobial usage, vaccination, global trade and travel (Daszak, 2012; Karesh et al., 2012; Morse et al., 2012).

Pandemics are generally viral in cause. This is thought to be due to their high mutation rate, which is particularly true for RNA viruses such as influenza where high nucleotide substitution and poor proof reading leads to the accumulation of errors in newly synthesized RNA strands. Influenza can also undergo re-assortment during mixed infection. These factors can result in divergence of surface antigens, such as haemagglutinin (HA) and neuraminidase (NA), producing strains not recognized by the human immune system and not covered by extant vaccines (Holland et al., 1982; Webster et al., 1992; Chen and Holmes, 2006; Hampson and Mackenzie, 2006; Jones et al., 2008; Taubenberger and Morens, 2008; Dormitzer et al., 2011; Morse et al., 2012). For instance, influenza A is now known to have 18 subtypes of HA and 11 subtypes of NA (Li et al., 2012; Tong et al., 2012; Wu et al., 2014). This high mutation rate and the emergence of new strains can also make vaccine development and policy difficult to plan and carry out. Due to viral antigenic shift, yearly influenza vaccines are required so the population is sufficiently protected by the vaccine, however, vaccine composition is determined ~8 months in advance of administration. This lag may allow new strains to emerge or for antigenic drift to result in a poor match between vaccine and the circulating strain of influenza. Furthermore as seen in the 2009 influenza pandemic, governments and public health departments face considerable difficulties in the production and distribution of vaccines when faced with sudden or unexpected outbreaks of newly emerged strains (Houser and Subbarao, 2015).

A common complication of respiratory viral disease can be secondary bacterial infection. Noting this association is important as it has clear implications for global health, principally because bacterial co/secondary infection is known lead to increased morbidity (Smith and McCullers, 2014). Co/secondary bacterial infection, as the name suggests, is a bacterial infection that occurs during or after an infection from another pathogen, commonly viruses. A number of viral infections (including infection from influenza virus, respiratory syncytial virus, parainfluenza virus and human metapneumovirus) can be complicated by co/secondary infection by a variety of bacteria including *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus*. This association leads to an increased severity of disease and sequela such as pneumonia (Smith and McCullers, 2014). In this review we dwell on

influenza pandemics since the late 1800's, focussing on the associations and complications that arise from secondary bacterial infections.

## INFLUENZA

Influenza viruses are important zoonotic pathogens as they are highly contagious and one of the most prevalent causes of respiratory infection. Worldwide annual epidemics reportedly cause up to five million cases of severe illness, which result in 250,000–500,000 deaths per year. The majority of deaths caused by influenza occur in young children and people over 65 (World Health Organization, 2016). Reports suggest that each year up to 20% of the United States population may be infected by influenza (Sullivan et al., 1993; Biggerstaff et al., 2014). The virus spreads easily from person to person via aerosol droplets (Hilleman, 2002; Taubenberger and Morens, 2008) and replicates in the upper and lower respiratory tract (Taubenberger and Morens, 2008). Commonly, in non-tropical regions, annual influenza epidemics occur during late autumn and winter. Although less frequent, tropical regions too suffer influenza epidemics, these generally coinciding with the rainy season (Cox and Subbarao, 2000; Biggerstaff et al., 2014).

There are three types of influenza virus, types A, B, and C, each differing in host range and pathogenicity (Taubenberger and Morens, 2008). Type A has been isolated from humans, avian, swine, horses, mink, dogs, seals, and ferrets (Jakeman et al., 1994; Taubenberger and Morens, 2008; Parrish et al., 2015), whilst type B has been isolated from humans, seals (Osterhaus et al., 2000) and ferrets (Jakeman et al., 1994), and type C from humans (Matsuzaki et al., 2002), swine and dogs (Yozubashi et al., 1996). Influenza A and B virions contain several structural antigens and three antigenic surface proteins; HA, NA, and M2/BM2 ion channels (Webster et al., 1992; Hampson and Mackenzie, 2006; Racaniello, 2009; Dormitzer et al., 2011). Influenza virus C only expresses one antigenic surface protein, haemagglutinin-esterase-fusion (HEF), and thus stimulates a lesser immune reaction than types A or B (Taubenberger and Morens, 2008; Racaniello, 2009). Influenza A is the fastest to evolve, at a rate 2–3 times faster than B, whilst C is the slowest (Yamashita et al., 1988). Antigenic drift allows the influenza virus to escape immunity acquired through previous exposure or vaccination; thus influenza A causes more epidemics and pandemics than either influenza B or C (Hampson and Mackenzie, 2006; Taubenberger and Morens, 2008). Whilst influenza B causes periodic/yearly epidemics but not pandemics, influenza C viruses only cause relatively infrequent mild respiratory problems (Taubenberger and Morens, 2008). Throughout the past 300 years there have been 12 pandemics caused by influenza A; the most infamous being the 1918 'Spanish flu' pandemic (Taubenberger and Morens, 2008). In the years between 1933 and 1957 there were nine influenza A (H1N1) epidemics and five influenza B epidemics. The worst of all these epidemics was the 1935–1936 influenza B epidemic that resulted in at least 55,000 deaths.

This was closely followed by the 1943–1944 influenza A (H1N1) epidemic which caused 53,000 deaths (Glezen, 1996). Evidently, although influenza B doesn't cause pandemics, it is still a cause for concern.

During an infection influenza virions attach to and enter host epithelial cells by the binding of viral HA to sialic acid on the host cell which instigates endocytosis and the movement of the virion into the cell within an endosome. The virus then uses/hijacks the host cells 'machinery' to replicate and transcribe viral RNA and produce more viral components (Samji, 2009). Progeny virions bud from the host cell, using the host cell membrane as a viral envelope, and go on to infect neighboring host cells (Nayak et al., 2009). As influenza infection develops the virus causes cell damage and death within the host's airways and up-regulates the production of toxins, causing further destruction. Influenza cytotoxins for example causes necrosis of host cells (Conenello and Palese, 2007; Iverson et al., 2011). Influenza infection, particularly pandemic influenza infection, is known to generate an increased inflammation response within the host, as the body works to rapidly deliver immune cells to the site of infection. This inflammation is a response to the expression of cytokines and chemokines (de Jong et al., 2006; Kash et al., 2006; Kobasa et al., 2007; Rock and Kono, 2008). Virally induced decreased mucociliary activity, the dysfunction of immune cells and the reduction of phagocytosis reduces clearance of the virus from the host airways and the host's ability to fight the virus (Brundage, 2006; Wu et al., 2011; Cauley and Vella, 2015). In an attempt to limit and control infection, the host immune system kills infected host cells. It does this in several ways, including; the production of perforin by Natural Killer (NK) cells which creates lesions/pores in cell membranes resulting in the induction of apoptosis, apoptosis from tumour necrosis factor (TNF) and FasL and the production of reactive oxygen species from macrophages and neutrophils causing oxidation of cellular lipids, proteins and DNA resulting in cell dysfunction and death (Topham and Hewitt, 2009; Kash et al., 2014; Kash and Taubenberger, 2015). Of course viral infection and/or interference with host processes can cause and direct the pathway of cell death, as is the case for necrosis. Host cell death, whether apoptosis, necrosis or pyroptosis, impacts on the severity and outcome of influenza disease in a variety of ways. Virally induced death of immune cells assist in the evasion of host defenses and hinders the clearance of the virus promoting the development of infection. Studies have shown a 90% reduction of alveolar macrophages in mice within a week of influenza infection, and evidence of necrosis in the remaining macrophages (Robinson et al., 2015). Necrosis and pyroptosis are pro-inflammatory due to their role in the release of cytokines. These cell death pathways allow for the rapid release of intracellular contents, including any viral components, from the infected host cell promoting host inflammatory responses and the formation of a cytokine storm which causes host tissue damage (Cundell et al., 1995; Rock and Kono, 2008; Lamkanfi and Dixit, 2010; Cauley and Vella, 2015). Furthermore infection with some influenza subtypes, for instance H1N1 and H5N1, typically result in lymphopenia, a state of abnormally low levels of lymphocytes, which has been associated with higher viral load.

de Jong et al. (2006) found influenza infection caused lower levels of cytotoxic T cell lymphocytes, which would therefore negatively affect acquired immunity (de Jong et al., 2006; Cunha et al., 2009). Where lymphopenia occurs, studies have shown a corresponding increase in macrophages. Supporting the evidence for the increase in macrophages is the significant increase in IP-10 (a chemokine secreted in response to gamma interferon (IFN $\gamma$ ) which activates macrophages), Interleukin-8 (IL-8, a chemokine which is produced by macrophages), IL-6 (in this case, a pro-inflammatory cytokine secreted by macrophages), and MCP-1 (a chemokine that recruits monocytes, a type of leukocyte that can differentiate into macrophages) (de Jong et al., 2006; Kobasa et al., 2007).

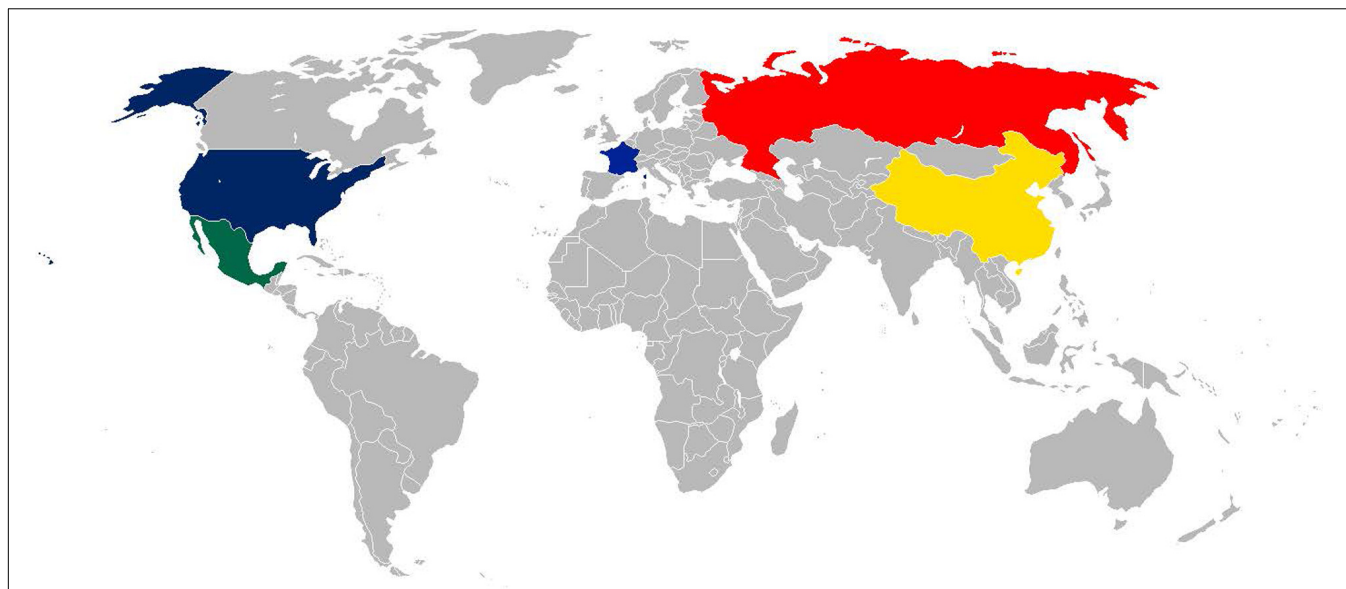
## INFLUENZA PANDEMICS SINCE THE LATE 1800's

Influenza pandemics, generally characterized by the emergence of a novel influenza A against which little or no immunity exists within the global populace, are a cause of high mortality and morbidity and are a major financial burden (Glezen, 1996). Since the 1800's these pandemics have arisen from a number of countries, spreading across the globe (**Figure 1**). Detailed below and in **Table 1** we have sought to describe some of the most significant influenza pandemics since the late 1800's to highlight the potential impact of influenza with respect to associations with bacterial infection.

## BACTERIAL CO-INFECTION AND SECONDARY INFECTIONS

Laboratory, clinical and epidemiological research has made it abundantly clear that bacterial co/secondary infection can significantly increase the morbidity and mortality of viral infections (Gupta et al., 2008). Up to 75% of those infected with influenza that go on to acquire pneumonia, are confirmed to have bacterial co-infection (Zambon, 2001). Bacterial co/secondary infection of influenza infection appears to occur frequently. Studies have shown that up to 65% of laboratory confirmed cases of influenza infection exhibited bacterial co/secondary infection, although Klein et al. (2016) state that in the majority of the research included in their meta-analysis this figure ranged between 11 and 35%. In the setting of an influenza epidemic or pandemic bacterial co/secondary infection can have devastating consequences, particularly in at-risk groups such as the immunocompromised/immunosuppressed. Immunosuppression is associated with more severe morbidity and a much higher risk of mortality from co/secondary bacterial infection (Rice et al., 2012). During the 2009 Swine influenza pandemic, there was an increase in hospital pneumonia cases as a result of secondary bacterial pneumonia, which was identified in 29–55% of mortalities (Centers for Disease Control and Prevention, 2009; Gill et al., 2010; Weinberger et al., 2012).





**FIGURE 1 |** World Map showing countries confirmed and suspected of being the origin of influenza pandemics. Blue – The origin of the 1918 Spanish is still unclear, although various papers suggest the United States (New York) or France as the origin; yellow – China the origin of the 1957 Asian flu pandemic; Hong Kong, the origin of the 1968 Hong Kong pandemic; red – Russia, the origin of the 1889 and 1977 Russian flu pandemics; green – Mexico, the origin of the 2009 Swine flu pandemic.

## PATHOBIONTS ASSOCIATED WITH CO/SECONDARY BACTERIAL INFECTION

The upper respiratory tract has been shown to host a diverse microbiota, within which a number of bacterial pathobionts may be found, i.e., those bacterial species that can be pathogenic yet also harmlessly carried (Hooper et al., 2012; Cauley and Vella, 2015). *Legionella pneumophila* (Rizzo et al., 2010), *Streptococcus pyogenes* (Chertow and Memoli, 2013), *Neisseria meningitidis*, *Moraxella catarrhalis*, *S. pneumoniae*, *H. influenzae*, *S. aureus* (Dela Cruz and Wunderink, 2017), *Pseudomonas aeruginosa* as well as a number of other *Streptococcus* and *Staphylococcus* spp. (Yang et al., 2016a) have all been associated with co-infection of influenza. However, *S. pneumoniae*, *H. influenzae*, and *S. aureus* are the most commonly reported bacteria associated with co/secondary infections during influenza pandemics since the late 1800's.

### ***Streptococcus pneumoniae***

*Streptococcus pneumoniae* is the most common bacteria found in viral secondary bacterial infections, and is particularly associated with causing high mortality and morbidity during influenza epidemics and pandemics (Brundage, 2006; Joseph et al., 2013). *S. pneumoniae* is a Gram-positive diplococci and is the most common cause of community-acquired pneumonia and invasive disease, i.e., sepsis and meningitis worldwide, as well as less severe acute disease such as otitis media (Bridy-Pappas et al., 2005; McCullers et al., 2010). *S. pneumoniae* is grouped into >97 immunologically distinctive serotypes based on a polysaccharide capsule (Bentley et al., 2006; Park et al., 2007; Jin et al., 2009;

Calix and Nahm, 2010; Calix et al., 2012). A burden to public health in its' own right, the WHO has reported that diseases caused by *S. pneumoniae* resulted in approximately 826,000 deaths in 2000 alone (Pittet and Posfay-Barbe, 2012). A more recent study shows that there are 4 million cases of disease caused by *S. pneumoniae* and 22,000 deaths annually in the United States (Huang et al., 2011). The current public health impact of *S. pneumoniae* infection is reduced by vaccine policies, with, for example, PCV-13 and PPV-23 being used for children and adults, respectively, in the United Kingdom (Pittet and Posfay-Barbe, 2012).

Many studies have shown that influenza infection facilitates the acquisition, colonization and development of disease from *S. pneumoniae* in people of all ages (Shrestha et al., 2013; Grijalva et al., 2014; Siegel et al., 2014). This is partly due to *S. pneumoniae*'s ability to catabolise sialic acid which is released from host cells and mucus by influenza's NA. Influenza infection also results in increased mucus production, further increasing the amount of metabolite available for *S. pneumoniae*. The NA produced by *S. pneumoniae* also assists in the release of sialic acid (Siegel et al., 2014). Mouse models support the concept that influenza facilitates the development of disease from *S. pneumoniae*; they have provided evidence that influenza infection enhances secondary *S. pneumoniae* pneumonia (McCullers and Rehg, 2002; McCullers and Bartmess, 2003). Wu et al. (2011), showed that co-infection of a virus and a bacterium can either occur from mixed viral bacterial infection, or from a viral infection being sequentially followed by a bacterial infection. Sequential bacterial infection normally occurs within a 7-day period of the viral infection. Influenza infections and successive *S. pneumoniae* infections result in a time and dose

**TABLE 1 |** Details of significant influenza pandemics since the late 1800's.

Name of pandemic	Year	Strain	Disease burden	Additional information
Russian influenza pandemic	1889	A(H2N2)	132,000 deaths in England, Wales, and Ireland alone (Smith, 1995).	The 1889 'Russian Flu' as the name suggests started in Russia and spread across Europe reaching North America in 1890. In only 4 months the infection had spread across all of Europe and the United States. The United Kingdom encountered four waves of disease and it is thought that at least one third of the adult population in England and Ireland suffered at least one bout of disease (Smith, 1995; Valleron et al., 2010).
Spanish influenza pandemic	1918	A(H1N1)	Caused 40–50 million deaths worldwide. India alone suffered 7 million deaths (Potter, 2001; Hilleman, 2002; Brundage, 2006; Michaelis et al., 2009).	Considered the most devastating influenza pandemic ever recorded, infecting 50% of the world's population. The origin of this pandemic is unclear as it appeared in North America, Asia, and Europe at roughly the same time (Taubenberger et al., 2001; Hilleman, 2002). Reports of disease and mortality were initially suppressed in many countries, included the United Kingdom, France and the United States, to ensure wartime efforts and morale weren't negatively affected. In Spain the press were able to print freely, meaning the first publicized cases were reported from Spain facilitating the nickname 'the Spanish flu' (Johnson, 2016; Peckham, 2016). In contrast to its name, it has been suggested that the pandemic started in France/mainland Europe and that it reached Spain from France (Reid et al., 2001; Trilla et al., 2008) although more recent papers suggest New York as the origin due to evidence of a pre-pandemic wave of the H1N1 virus (Olson et al., 2005). What is remarkable is how far the pandemic spread; the pandemic reached as far as the Alaskan wilderness to remote Pacific islands (Burnet and Clark, 1942; Taubenberger et al., 2001). The pandemic experienced a couple of waves; the first of which was relatively mild. The second wave, however, was far more lethal (Hilleman, 2002). The first outbreaks were reported in military camps as males responded to the call for troops in the spring and summer of 1918. A period of dormancy was then recorded toward the end of summer in America, but this was short lived as transmission picked up as schools reopened in September after the summer holidays (Glezen, 1996).
Asian influenza pandemic	1957–1958	A(H2N2)	Although global death toll estimates vary [between 1.5 million (Gatherer, 2009) and 2–4 million (Michaelis et al., 2009)], the death toll in the United States is accurately reported to have been 69, 800 (Klimov et al., 1999; Hilleman, 2002).	The pandemic affected 40–50% of people worldwide (Potter, 2001), however, resulted in lot less mortality than the previous pandemic. This Asian influenza pandemic started in March 1957 in Southern China, where pigs, ducks, and humans live together closely. It reached Hong Kong in April, and then spread to Singapore, Taiwan, and Japan (Fukumi, 1959; Potter, 2001; Hilleman, 2002). The pandemic reached India, Australia, and Indonesia by May, Pakistan, Europe, North America, and the Middle East by June, South Africa, South America, New Zealand, and the Pacific Islands by July, and Central, West and East Africa, Eastern Europe, and the Caribbean by August (Dunn, 1958; Payne, 1958; Potter, 2001).
Hong Kong influenza pandemic	1968–1969	A(H3N2)	1–2 million people died worldwide (Michaelis et al., 2009). Overall 33,800 people died in the United States (Klimov et al., 1999) and England and Wales saw a 55% increase in respiratory deaths in 1969 (Tillett et al., 1983).	The 1968 Hong Kong pandemic started in July 1968 in Hong Kong and spread to the Southern hemisphere by June 1969 (Biggerstaff et al., 2014). The H3N2 virus was isolated and identified too late in the pandemic for vaccine intervention (Nakajima et al., 1978; Hilleman, 2002) so it was fortunate that in most countries, apart from the United States, the disease was mild (Cockburn et al., 1969). There are several proposed reasons for the reduced mortality of this compared to the Asian Flu. Firstly the N2 was seen in the Asian Flu so may have contributed some cross-reactive immunity to this H3N2 strain (Glezen, 1996). Although antibodies to NA do not prevent infection, they help to reduce the amount of newly formed virus released from infected cells (Couch et al., 1974; Glezen, 1996). Secondly, during the initial wave of this pandemic, the number of cases started to grow exponentially in December, at this point the school Christmas holidays began; it has been speculated that this removed an important susceptible population (Glezen, 1996).
Russian Flu influenza pandemic	1977–1978	A(H1N1)	Approximately 700,000 deaths globally (Michaelis et al., 2009).	This pandemic was caused by a reappearance of H1N1, identical to that of the Spanish flu virus (Michaelis et al., 2009). The disease mainly affected those born after the late 1950's, so those who had not been exposed to the pandemic H1N1 strain that had circulated previously (Hilleman, 2002).
Swine influenza pandemic	2009	A(H1N1)	By the end of the pandemic it is thought that there were 284,000 deaths worldwide (Chertow and Memoli, 2013).	In early 2009, an influenza A H1N1 virus outbreak was initially identified in Mexico and then the United States (Michaelis et al., 2009). In June 2009 the WHO declared the outbreak a pandemic. Within 4 weeks the outbreak had spread to 41 countries, resulting in 11,034 confirmed cases and 85 deaths (Michaelis et al., 2009; Wang and Palese, 2009). Disease/symptoms were generally mild (Peiris et al., 2009) however, complications of the disease did result in hospitalization, particularly in at risk groups (Wang and Palese, 2009).

(Continued)

TABLE 1 | Continued

Name of pandemic	Year	Strain	Disease burden	Additional information
				Unlike other pandemics and yearly epidemics, during this pandemic it was predominantly children and young adults that were affected, particularly those aged 12–22 (Gill et al., 2010). Overall this pandemic was relatively mild. It is thought that morbidity and mortality rates were reduced due to three main factors. Firstly, the quick responses of various governments in terms of school closures helped reduce the spread of the virus. Thousands of schools were shut worldwide, including the United States and Mexico. Japan alone closed almost 2000 schools (Wang and Palese, 2009; Jackson et al., 2014). Secondly, influenza A H1N1 strains have been circulating amongst the human population for decades, therefore prior exposure could have provided some degree of immunity against the 2009 pandemic strain. Lastly an important pathogenicity factor, PB1-F2, was not present making the strain milder than those present in previous pandemics (Wang and Palese, 2009).

dependent change in the host dendritic cells which produces enhanced inflammation. Berendt et al. (1975) inoculated squirrel monkeys with either influenza A, *S. pneumoniae* or influenza A and *S. pneumoniae*. Influenza alone caused minor illness such as mild tracheitis, with symptoms such as sneezing, coughing and fever (although some did develop bronchopneumonia) and had a 100% survival rate. *S. pneumoniae* again caused minor illness with a 100% survival rate. Co-infection of influenza A with *S. pneumoniae* resulted in severe morbidity with a 75% death rate within 40 h, clear evidence of the consequences of co/secondary bacterial infection (Berendt et al., 1975). These findings are reflected in several other studies, with some even showing that co-infection may assist in the spreading of *S. pneumoniae* infection to the lower respiratory tract (Takase et al., 1999; Seki et al., 2004).

An additional mouse model of infection provided comparable results whilst comparing the effect of different *S. pneumoniae* serotypes on co-infection (Sharma-Chawla et al., 2016). More cases of pneumonia and bacteraemia were observed in mice infected with both influenza A and *S. pneumoniae* than in mice infected with these pathogens individually. This was the case for all *S. pneumoniae* serotypes tested. More virulent pneumococcal serotypes caused a greater burden of disease in both the co-infected mice and those infected with *S. pneumoniae* alone. The highly invasive pneumococcal serotype 4 caused pneumonia in 58% of mice and bacteraemia in 21% in a single infection model. When co-infecting with influenza these figures increased to 100 and 90% for pneumonia and bacteraemia, respectively. Mortality rates increased from 0% for individual infection to 79% during co-infection. In comparison, individual infection with a carrier strain (of lower invasive potential) of serotype 19F, caused pneumonia in 91% of cases and bacteraemia in 0%. When co-infecting with influenza and 19F these figures increased to 100 and 33%. Mortality rose from 0% during individual infection to 63% during co-infection (Sharma-Chawla et al., 2016).

Pneumococcal vaccination has shown to ameliorate the risk of secondary bacterial pneumonia. During a vaccine efficacy study, the incidence of pneumonia in those with influenza reduced by 45% in groups vaccinated against *S. pneumoniae* (Madhi et al., 2007). However, whilst vaccine implementation has successfully reduced pneumococcal disease in a number of countries, lower levels of vaccine implementation in low and middle income countries coupled with fractional serotype

coverage and increasing levels of antibiotic resistance, means the specter of influenza pandemic associated *S. pneumoniae* secondary infection remains a significant risk to global health.

### ***Haemophilus influenzae***

*Haemophilus influenzae* is another bacteria commonly found to co/secondarily infect viral infection, and has been associated with the complication of disease during influenza pandemics (Abrahams et al., 1919; Spooner et al., 1919; Brundage, 2006; Palacios et al., 2009). It is a Gram-negative fastidious coccobacillus. Typeable strains have a polysaccharide capsule and are categorized into six serotypes (A–F). *H. influenzae* serotype B was a major cause of invasive disease (Wenger, 1998; Murphy, 2003; Chin et al., 2005; Brouwer et al., 2010) although widespread implementation of the Hib vaccine has significantly reduced the burden of disease (Rosenstein and Perkins, 2000). Those *H. influenzae* that lack a capsule, denoted non-typeable *H. influenzae* (NTHi), remain a significant cause of bacterial meningitis, otitis media and exacerbations of chronic lung disease such as COPD worldwide (Langereis and de Jonge, 2015).

Various studies have shown the impact when *H. influenzae* co/secondarily infects with influenza, and some suggest a level of synergism. The effect of influenza and *H. influenzae* co-infection verses individual infection of both pathogens is tellingly different; Shope found that co-infection resulted in severe disease or death when on their own *H. influenzae* and influenza only caused mild infection or disease (Shope, 1931). More recently, Lee E.H. et al. (2010) undertook a similar study which provided comparable results and evidence that influenza and *H. influenzae* co-infection produces more epithelial cell destruction than single infection with either pathogen (Lee L.N. et al., 2010). Furthermore, they found individual infection caused mild bronchiolitis within 4 days of initial infection, from which the host lung was able to recover. Conversely, co-infection caused bronchial necrosis, bronchial inflammation and bronchitis within the same time period or less, and led to further complication such as epithelial erosion (Lee L.N. et al., 2010). It is now commonly accepted that co-infection results in more severe morbidity and poorer clinical outcome than infection of influenza or *H. influenzae* alone.

Further support of the impact of co-infection comes from Michaels et al. (1977), who dosed two groups of rats intranasally with *H. influenzae* with the intention of giving them meningitis. One group of rats were naive and the other had previously been

dosed with influenza. In both groups ~50% of the rats acquired meningitis, however, the naïve rats required a 100-fold larger dose of *H. influenzae* (Michaels et al., 1977).

As is the case for many bacterial and viral co-infections, mortality from *H. influenzae* and influenza co-infection is highly dependent on the timing of the introduction of the secondary microbe as well as density of bacterial colonization. Studies have shown that when influenza virus and *H. influenzae* are introduced at the same time there is no synergistic relationship. When *H. influenzae* is introduced 7 or more days after influenza there is again no synergistic relationship; however, high lethality is exhibited when *H. influenzae* and influenza are introduced 3 or 4 days apart (Lee L.N. et al., 2010).

## ***Staphylococcus aureus***

*Staphylococcus aureus* is a Gram-positive cocci that has been found to complicate influenza infection; increasingly so in more recent years/pandemics (Hers et al., 1958; Palacios et al., 2009). *S. aureus* is transiently carried in the nose of 30% of the population, whilst 20% of the population have persistent nasal colonization (Wertheim et al., 2005). Like *H. influenzae* and *S. pneumoniae*, *S. aureus* is an opportunistic pathogen and a major cause of bacteraemia (Wertheim et al., 2005; Tong et al., 2015). It is also a common cause of pneumonia (Kollef et al., 2005); specifically necrotising pneumonia that is caused by community acquired Methicillin-resistant *Staphylococcus aureus* (MRSA) and has a 30% mortality rate (Murray et al., 2010). Necrotising pneumonia is highly associated with either the presence of Panton-Valentine leukocidin (PVL) or prior/co influenza infection (DeLeo and Musser, 2010). MRSA is a particularly problematic pathogen and concern for public health as it can be hard to treat due to its multidrug-resistant properties (Wu et al., 2005; Eom et al., 2014; Fishovitz et al., 2014).

Influenza infection has been shown to increase the adherence of *S. aureus* (as well as *H. influenzae* and *S. pneumoniae*) to host pharyngeal cells (Fainstein et al., 1980). In addition to this, mouse models have highlighted increased morbidity and mortality in mice that are pre-infected with influenza before they are exposed to *S. aureus* vs. those just exposed to *S. aureus*. Increased lung damage and bacterial density has also been shown (DeLeo and Musser, 2010; Lee M.H. et al., 2010; Iverson et al., 2011). Lee M.H. et al. (2010) showed that mice infected with low doses of influenza, low doses of *S. aureus* and high doses of *S. aureus* were able to survive. Those infected with high doses of influenza died within 4–7 days; however, all mice infected with a high dose of influenza and then a high dose *S. aureus* died within 2 days of bacterial exposure, showing how death can be accelerated by co-infection. When mice were infected with a low dose of influenza and then a high dose *S. aureus* they died at 7 days. The fact that the mice survived low influenza infection on its own, but could not survive co-infection with *S. aureus* shows the lethality of such co/secondary bacterial infection (Lee M.H. et al., 2010).

In an act of synergism, *S. aureus* infection may actually assist influenza infection by increasing the infectivity of influenza; when the virion is being moved into the host cell within an endosome the low pH in the endosome causes a conformational change to the HA [HA<sub>(0)</sub>] allowing it to be cleaved by host

proteases into two subunits [HA<sub>(1)</sub> and HA<sub>(2)</sub>]. This cleaving ‘activates’ the HA, mediating fusion between the virus and endosome membrane, ready for the opening of the M2 ion channel so the vRNP (viral ribonucleoproteins) can be released into the host cell where the viral RNA is replicated and transcribed. Several strains of *S. aureus* produce proteases that cleaves influenza HA; the more protease that is available, the more HA can be cleaved meaning more vRNP can get into host cells meaning overall more progeny virions (Tashiro et al., 1987; Steinhauer, 1999; Samji, 2009). This aspect contributes to the increased severity of disease caused by co-infection verses individual influenza infection. And although not all strains of *S. aureus* produce proteases that cleave influenza HA, the proteases they do produce indirectly enhance morbidity by causing host inflammatory responses which result in the production of host enzymes that are capable of cleaving HA (Tashiro et al., 1987).

## **HISTORICAL EVIDENCE OF CO/SECONDARY BACTERIAL INFECTION DURING MAJOR INFLUENZA PANDEMICS**

### **1918 Spanish Influenza Pandemic**

The 1918 influenza pandemic was a result of influenza strain A (H1N1). It is considered the most devastating influenza pandemic ever recorded, infecting 50% of the world’s population and resulting in approximately 40–50 million deaths worldwide. India alone suffered 7 million deaths (Potter, 2001; Hilleman, 2002; Brundage, 2006; Michaelis et al., 2009). The main groups of individuals affected by this pandemic were those aged 20–40 years old, in addition to infants and those over 65. Ordinarily only young children and the elderly are the age groups most at risk from influenza, showing how distinctive pandemic strains can be (Potter, 2001). It is suggested that war time efforts meant that influenza easily spread through military camps, allowing the 20–40 years old age range to be more at risk than usual.

There are many published examples of co/secondary bacterial infections during the 1918 influenza pandemic, and pneumonia as a consequence of bacterial infection is estimated to have occurred in up to 95% of deaths during this pandemic (Morens et al., 2008). A majority of those deaths due to secondary *S. pneumoniae* infection (Brundage and Shanks, 2008; Morens et al., 2008). Many of the examples that detail co/secondary bacterial infection come from outbreaks within army camps. Within a 1-month period in 1918 at the military Camp Devens, a quarter of all troops were diagnosed with influenza. Of those infected, 17% developed pneumonia, of which 35% of cases were fatal. Out of 37 autopsies performed, 43% were positive for pure growth of *H. influenzae* in at least one lobe of the lung. Blood culture revealed 65% had *S. pneumoniae*, 2.5% had *H. influenzae* and 1.3% had *S. aureus* (Spooner et al., 1919; Brundage, 2006). This pattern of invasive bacterial co/secondary infection has also been documented for several other camps during the same year, including Camp Logan. Here 2,487 influenza-associated



hospitalizations were recorded, 17% acquired pneumonia with 4% of these cases being fatal. Post-mortems found *S. pneumoniae* in the lungs of 44% and heart blood of 33% (Hall et al., 1918; Brundage, 2006). At Camp Jackson, 17% of influenza cases progressed to pneumonia with a further 31% of pneumonia cases proving fatal. Autopsies found *S. pneumoniae* to be the bacterial co-infection most associated with pneumonia, however, 155 of 312 lung cultures were positive for *S. aureus* (Michael, 1942; Brundage, 2006). At Camp Custer, 21% of influenza cases progressed to pneumonia, of which 28% died. Sputum cultures proved the presence of *S. pneumoniae* in 26% of cases. Further investigation found 28% of lung and blood cultures were positive for *S. pneumoniae*, again acting as supporting evidence of the invasive potential of such co-infections (Blanton and Irons, 1918; Brundage, 2006). Camp Fremont experienced 2418 hospitalizations, 17% had pneumonia of which 36% were fatal. Nasopharyngeal and sputum samples from 158 pneumonia cases found *S. pneumoniae* in 41% of cases, *H. influenzae* in 38%, and other *Streptococcus* spp. in 29% (Brem et al., 1918).

Further lung tissue from fatalities of this pandemic were re-examined in 1919; *S. pyogenes longus* was found in 36% of cases, *S. pneumoniae* in 29% of cases and *H. influenzae* in 25% (Abrahams et al., 1919; Brundage, 2006). Additional post-mortems of lung tissue suggest that at least 90% of samples showed evidence of bacterial infection (Oxford et al., 2002; Morens et al., 2008; Chien et al., 2009). Overall 95% of deaths were due to co/secondary bacterial pneumonia (Opie et al., 1921; Morens et al., 2008).

Co-infection had also been reported as an issue prior to the official start of the pandemic. Influenza with secondary bacterial infection of *S. pneumoniae* (and other *Streptococcus* sp.), *H. influenzae* and/or *Staphylococcus* sp. was associated with major outbreaks of purulent bronchitis in 1916 and 1917 (Brundage, 2006; Joseph et al., 2013). Indeed in 1916–1917 British, Australian, Canadian, and American armed forces in England and France experienced an epidemic of purulent bronchitis. Out of 20 tested sputum specimens from a British army camp based in north France, 90% presented with *H. influenzae*, 65% presented with *S. pneumoniae*, 25% with other *Streptococcus* spp. and 15% with *Staphylococcus* spp. Out of the specimens positive for *H. influenzae*, many exhibited simultaneous *H. influenzae* and *S. pneumoniae* co-infection; with *H. influenzae* identified as the primary bacterial infector. *S. pneumoniae* infection first presented with low virulence, however, pathogenesis soon worsened, it has been suggested, as result of the symbiotic growth with *H. influenzae* (Brundage, 2006; Dennis Shanks et al., 2012). Of course it is known that there is a positive association between the colonization of *H. influenzae* and *S. pneumoniae*, and colonization is a prerequisite for disease, so the presence of such co-infection fits with current knowledge (Jacoby et al., 2007; Abdullahi et al., 2008).

## 1957 Asian Influenza Pandemic

This pandemic affected 40–50% of people worldwide. The cause was influenza strain A (H2N2) (Potter, 2001). Although global death toll estimates vary [between 1.5 million (Gatherer, 2009)

and 2–4 million (Michaelis et al., 2009)], the death toll in the United States is accurately reported to have been 69,800 (Klimov et al., 1999; Hilleman, 2002). Post-mortem cultures show evidence of bacterial infection in up to 80% of all severe and fatal cases (Hers et al., 1958; Morens et al., 2008; Gill et al., 2010).

During this pandemic the United States, and many other countries, experienced an increase in hospitalization rates. A majority were due to pneumonia, predominantly caused by *S. pneumoniae*, *H. influenzae*, and *S. aureus* infection (Petersdorf et al., 1959). There are similar documented reports from the Netherlands; of the 148 deaths presumed to be from the Asian pandemic influenza strain that were examined fully, 75% presented with bacterial pneumonia of which 15% were positive for *S. pneumoniae* and 59% were positive for *S. aureus* (Hers et al., 1958).

Robertson et al. (1958) unveiled similar findings when investigating the hospitalization of 140 people suffering pneumonia at Sheffield City General Hospital in 1957. A majority showed evidence of influenza A infection; 27% of those had co/secondary infection of *S. aureus* (which had a 47% death rate), 15% *S. pneumoniae* and 4% *H. influenzae*, although this is likely to be an underestimation as many patients had already started taking antibiotics (Robertson et al., 1958).

## 1968–1969 Hong Kong Influenza Pandemic

Worldwide 1–2 million people died during this pandemic which was caused by the influenza strain A(H3N2) (Michaelis et al., 2009). Although this is a lower death toll than engendered in previous pandemics, it is still an awfully high number of deaths. Overall 33,800 people died in the United States (Klimov et al., 1999) and the pandemic cost 3.9 billion dollars (Hilleman, 2002). In 1969, England and Wales saw a 55% increase in respiratory deaths, of which co/secondary bacterial infection was shown to be a major contributor (Tillett et al., 1983).

Staphylococcal pneumonia in particular was a major source of complication to influenza infection. A hospital in Atlanta suffered a threefold increase in cases of Staphylococcal pneumonia during this pandemic. Staphylococcal infection caused 26% of pneumonia cases during this period, and a high correlation was recognized between influenza infection and bacterial pneumonia (Schwarzmann et al., 1971). In addition, out of 79 cases of fatal influenza with pneumonia complications, 16% had bacterial co-infection with *S. pneumoniae* (6%), *S. pyogenes* (5%), and *S. aureus* (1%) being the main causes. More than one of these bacteria were present in 4% of cases (Schwarzmann et al., 1971; Metersky et al., 2012).

Another health care facility in the United States, the Mayo Clinic in Minnesota, also found *S. aureus* to be a major cause of complication. Of 129 adults diagnosed with pandemic influenza, pneumonia was established in 16%, of which 40% of these cases (6% of all 129 influenza cases) were fatal. *S. aureus* or *P. aeruginosa* bacterial infection was present in 75% of all fatal cases, indicating bacterial co/secondary infection was a major determinant of severe disease and death (Lindsay et al., 1970).

In previous pandemics *S. pneumoniae* has been proposed as the major contributor of mortality and morbidity, however, during this 1968–1969 Hong Kong and the 1957 Asian influenza pandemic *S. aureus* clearly had a larger impact. This is possibly a reflection of increased antibiotic use and increased antibiotic resistance.

## 2009 Swine Influenza Pandemic

Within 4 weeks this outbreak of influenza A(H1N1) had spread to 41 countries resulting in 11,034 confirmed cases and 85 deaths (Michaelis et al., 2009; Wang and Palese, 2009). By the end of the pandemic it is thought that there were 284,000 deaths worldwide, with Mexico and the United States being most severely affected (Chertow and Memoli, 2013). Unlike other pandemics and yearly epidemics, during this pandemic it was predominantly children and young adults that were affected, particularly those aged 12–22 (Gill et al., 2010). Influenza A (H1N1) strains have been circulating amongst the human population for many years therefore this prior exposure could have provided many adults with some degree of immunity against the 2009 pandemic strain, particularly older groups who were more likely exposed during previous pandemics.

Surveillance by the New York City Department of Health and Mental Hygiene has shown that during the 2009 Swine Flu Pandemic almost 30% of the first 47 deaths showed invasive bacterial disease. *S. pneumoniae* was the most common causative agent identified (followed by *S. pyogenes*) (Lee E.H. et al., 2010). In the United Kingdom, of the 457 fatalities 68 were autopsied. Of these, 41% were shown to have complications associated with secondary bacterial infection, most commonly (25% of cases) due to *S. pneumoniae* (Lucas, 2010).

Further studies in the United States have reviewed 77 deaths during the period of May–August 2009 and found bacterial co-infection in almost 30% of cases; 46% of which were with *S. pneumoniae*, 9% with *S. aureus* and 1% with *H. influenzae* (Centers for Disease Control and Prevention, 2009). Studies based in Argentina produced similar evidence for the presence of bacterial infection, showing this wasn't just a localized trend. Palacios et al. (2009) examined nasopharyngeal swab samples from almost 200 cases of pandemic influenza. *H. influenzae* was found in 52%, *S. pneumoniae* was found in 31% and *S. aureus* in 18% of samples. Although not the most common bacteria found, *S. pneumoniae* was the most strongly associated with severe disease (Palacios et al., 2009).

Additional research in pediatric intensive care units in the United States, investigated 838 critically ill children who were infected with pandemic influenza. Within 72 h of admission to the intensive care unit 33% exhibited bacterial co-infection; in 26% of these cases *S. aureus* was identified as the cause (48% of which were MRSA), 5.5% were positive for *S. pneumoniae* and 5% were positive for *H. influenzae*. Bacteraemia was observed in 5% of admissions, for which *S. aureus* was the main cause (Randolph et al., 2011). This study highlights how quickly co/secondary bacterial infection can become invasive particularly in at risk groups such as young children or the elderly. A point of concern is that almost half of the *S. aureus* were MRSA, and therefore inherently resistant to multiple antibiotics.

In another study of vulnerable and critically ill children in a pediatric intensive care unit in the United States, 51% of those with influenza infection had bacterial co/secondary infection. Of these 35% presented with *S. aureus*, 18% *P. aeruginosa*, 18% *M. catarrhalis*, 9% NTHI, 6% *S. pneumoniae* and 6% Group A Streptococcus. Those with *S. aureus* showed more severe morbidity and were more likely to develop disseminated intravascular coagulation which leads to a compromised blood flow within body tissue and therefore tissue damage (Nguyen et al., 2012).

In a retrospective study of 50 patients who were infected during pandemic influenza, 28% showed co/secondary bacterial infection (Dhanao et al., 2011). *Mycoplasma pneumoniae* was found in 10%, making it the most common co/secondary infecting bacteria. This was followed by *S. aureus* found in 6%, *K. pneumoniae* and *S. pneumoniae* found in 4% and *M. catarrhalis*, *P. aeruginosa*, *S. pyogenes*, and *Streptococcus agalactiae* found in 2% of these patients (Dhanao et al., 2011).

*Moraxella catarrhalis* is a bacteria of increasing importance being now acknowledged as the third most common cause of otitis media (OM), after *S. pneumoniae* and *H. influenzae* (Bluestone, 1986; Faden et al., 1994; Kilpi et al., 2001; Dupont et al., 2010) and the second most common cause of exacerbations in COPD, accounting for up to 4 million exacerbations per year in the United States alone (Murphy et al., 2005). *M. catarrhalis* is a cause of pneumonia (Berg and Bartley, 1987; Hager et al., 1987; Marchant, 1990; Verduin et al., 2002) and invasive disease such as bacteraemia (Ioannidis et al., 1995) and meningitis (Newing and Christie, 1947), with bacteraemia being a common complication of pneumonia, particularly in adults (Collazos et al., 1992; Ioannidis et al., 1995). Although this review has focused on *S. pneumoniae*, *H. influenzae*, and *S. aureus*, it has cited other bacteria seen as a source of co-infection during the various pandemics described. In early influenza pandemics such as the 1918 Spanish pandemic, *M. catarrhalis* rarely appears to be a noted cause of co-infection. However, in the 2009 pandemic it is seen in up to 18% of cases (Nguyen et al., 2012). We have therefore considered the importance of this. Data produced toward the end of the 1970's and throughout the 1980's demonstrated *M. catarrhalis*' potential to cause disease, however, before this *M. catarrhalis* was considered a non-pathogenic harmless commensal (McNeely et al., 1976; Johnson et al., 1981; Onofrio et al., 1981; McLeod et al., 1983; Feder and Garibaldi, 1984; Hager et al., 1987; Catlin, 1990). Therefore there are two possibilities to consider; perhaps *M. catarrhalis* wasn't present in early pandemics as a cause of co-infection and has become more of an issue in recent years; possibly as a result of vaccines, i.e., Hib and PCV, reducing the disease burden of other bacteria such as *S. pneumoniae* and *H. influenzae*. Alternatively, we must consider that as *M. catarrhalis* was not considered a pathogen it was therefore missed or not commented upon prior to the 1980's. Retrospective studies may be able to address this. For example, autopsies from the 1918 pandemic were reviewed and it was found that *S. pneumoniae* was the most common co-infecter, followed by *S. hemolytic*, *S. aureus*, and *H. influenzae*. 'Other bacteria' were also

highlighted within which *M. catarrhalis* was grouped (Morens et al., 2008).

Another point of consideration are changes of methodology. Pre-1983 laboratories would only undertake bacterial culture, however, in 2009 more sensitive methodology, i.e., PCR were available and commonly used in laboratories worldwide. The use of sensitive methods such as PCR, may have increased the likelihood of *M. catarrhalis* being detected, and as a known respiratory pathogen it would have been tested for, where as previously it may not have been. Alternatively maybe PCR detects bacteria that may have been out grown/not shown on a culture plate?

In contrast to *S. pneumoniae* and *H. influenzae* little research has been undertaken looking at influenza and *M. catarrhalis* co-infection and the dynamics and mechanisms of such infection. This is therefore an area worthy of future research. *M. catarrhalis* has been highlighted as a frequent source of co-infection for influenza since the early 1980's (Klein et al., 2016). In the setting of a pandemic it may therefore have a major public health impact.

## FACTORS AFFECTING THE SEVERITY OF BACTERIAL CO/SECONDARY INFECTION

As discussed above, co/secondary bacterial infection can result in a deterioration of clinical condition with more severe disease. The severity of co/secondary infection depends on multiple factors such as the strain of virus and serotype/strain of bacteria, the lag between viral infection and bacterial exposure and density of bacterial colonization (Lee L.N. et al., 2010; McCullers et al., 2010; Smith and McCullers, 2014).

### Virally Enhanced Colonization and Attachment of Bacteria

It has become clear that influenza, as well as other upper respiratory tract viral infections, leads not only to a greater risk of infection from bacterial pathobionts but also an increased likelihood that an individual may become colonized with bacteria such as *S. pneumoniae*, *H. influenzae* and *S. aureus* (Hament et al., 1999; Hilleman, 2002). Plotkowski et al. (1986) found enhanced colonization and adherence of *S. pneumoniae* to the tracheal cells of mice when they were infected with influenza (Plotkowski et al., 1986). Other studies have intranasally inoculated ferrets with influenza, finding prior viral infection increases colonization and adherence of *S. aureus* (Sanford and Ramsay, 1987). Furthermore, poor disease outcome has been linked to lost lung repair function and loss of basal epithelial cells, including alveolar epithelial cells; which is associated with increased bacterial attachment and apoptosis (Kash et al., 2011). Wadowsky et al. (1995) conducted a study in which adult subjects were inoculated with influenza and then screened for bacterial colonization. After 6 days 15% of the subjects were heavily colonized by *S. pneumoniae* (Wadowsky et al., 1995). Additionally, the effect of viral prevention methods further supports the idea of viruses predisposing a host to secondary bacterial infection (Peltola

and McCullers, 2004; Lee et al., 2008). Studies have shown that influenza vaccination can reduce the occurrence of bacterial pneumonia (Fedson et al., 1993; Nichol et al., 1998).

## Viral Factors Implicated in Severity of Infection

Research shows that influenza A is the type most commonly associated with co/secondary bacterial infection and subtypes with NA2 traditionally result in more severe infection (Peltola et al., 2005). Although reported less, influenza B has also been associated with severe bacterial co/secondary infection (Finelli et al., 2008; Aebi et al., 2010). Various factors are known to impact the severity of viral infection, which in turn increases the likelihood of bacterial co/secondary infections; these include the type of HA and NA surface antigen. As mentioned previously, HA mediates virion binding to the host cell via sialic acid receptors. Binding is followed by endocytosis and the movement of the virion into the host cell within an endosome (Samji, 2009). HA binds to sialylated glycans found on the surface of human epithelial cells; traditionally seasonal influenza A virus binds to  $\alpha$ 2-6 sialylated glycans on cells in the upper respiratory tract whereas the highly pathogenic avian H5N1 strain binds to  $\alpha$ 2-3 sialylated glycans on type 2 pneumocytes lining lung alveoli (Shinya et al., 2006). Clearly the type of HA impacts on the site and development of infection. The low pH in the endosome causes a conformational change to the HA allowing it to be cleaved, an important step in penetrating into the host cell. Therefore HA and the availability of appropriate host proteases are determinants of infectivity (Steinhauer, 1999; Samji, 2009). Interestingly non-pathogenic and mammalian influenza HA undergoes cleavage outside of the host cell where as highly pathogenic strains are cleaved inside host cells (Steinhauer, 1999). Another example of how the type of HA can make a difference to infection, and therefore the impact of an epi- or pandemic, is that traditionally trypsin-like protease cleaves influenza HA; however, some HA types (i.e., types 5 and 7) have the ability to acquire insertional mutations at the cleavage site which changes their recognition site in such a way that specificity is broadened so more proteases are recognized (Kash and Taubenberger, 2015).

Neuraminidase enables the release of newly formed progeny virions; by hydrolysing the sialic acid and detaching it from the HA the virion becomes liberated from the host cell (Zambon, 2001). To be truly effective the NA must be complementary and share the same receptor specificity as HA, so if the viral HA binds to  $\alpha$ 2-3 sialic acid then the NA should hydrolyse  $\alpha$ 2-3 sialic acid (Baum and Paulson, 1991).

The production of viral toxins that impact host cell integrity is another important factor in the development of co/secondary bacterial infection. Influenza A virus can produce a viral cytotoxin PB1-F2 (Conenello and Palese, 2007; Iverson et al., 2011) which plays a role in increasing inflammation and therefore host cell damage and bacterial adherence, increasing mortality and morbidity (Lee et al., 2016). It also helps reduce bacterial clearance, increasing the occurrence and severity of co/secondary bacterial infection, by causing cell death in host monocytes (Conenello and Palese, 2007; Iverson et al., 2011).



## Molecular Co-pathogenesis

Following bacterial colonization, disease develops due to specific characteristics of viral infection that facilitate bacterial adhesion and penetration (Selinger et al., 1981). Influenza produces NA, which increases adhesion of some bacterial species by removing sialic acid to expose host cell receptors (McCullers and Bartmess, 2003; Peltola and McCullers, 2004). Alternatively some bacteria, i.e., group B Streptococci, contain sialic acid which allows for direct binding to the viral HA expressed by influenza infected host cells (Okamoto et al., 2003; Peltola and McCullers, 2004). Damaged host cells, whether damaged directly by the virus or by inflammation and immune cell responses, provide additional adhesion sites allowing for increased bacterial adhesion. For example the exposure of apical receptors like integrins permit the adhesion of bacteria such as *S. aureus* and *P. aeruginosa* (Sanford et al., 1978; Davison and Sanford, 1981; Bucior et al., 2012; Smith and McCullers, 2014). In response to viral infection, host inflammatory responses may cause an up-regulation in the expression of host receptor molecules and other molecules that bacteria can use as receptors (Hakansson et al., 1994; Peltola and McCullers, 2004). For example Cundell et al. (1995) showed an increased presentation of G-protein-coupled platelet-activating factor (PAF) receptor, which certain bacteria, i.e., *S. pneumoniae*, can utilize for cell attachment and colonization in endothelial cells (Cundell et al., 1995; van der Sluijs et al., 2010). In contrast it has been suggested that the PAF receptor does not affect initial bacterial adherence and colonization but is more involved with assisting bacterial transition/spread into the blood and thus the development of invasive disease (McCullers et al., 2008).

Influenza infection appears to prime the host airways for bacterial infection, whilst modifying and impairing immune responses in a number of ways (Joseph et al., 2013). Viral induced immunosuppression can allow for a bacterial super infection, as host immune responses can be suppressed when immunologic cells are impaired during influenza infection and immune cell dysfunction can reduce the host's ability to fight bacteria (Peltola and McCullers, 2004; Brundage, 2006; Wu et al., 2011). Many studies involving animal models have shown that influenza infection increases and prolongs bacterial growth, due to reduced macrophage accumulation and decreased bacterial clearance due to reduced phagocytic activity (Kleinerman et al., 1976; Wyde et al., 1989; Sun and Metzger, 2008). Additionally it has recently been shown that *S. pneumoniae* and influenza co-infection results in a reduction in the number of local alveolar macrophages, this due to increased death of these macrophages by apoptosis and necrosis (Sharma-Chawla et al., 2016). This reduction is likely to hinder bacterial clearance, hence the increased bacterial load found during co-infection during this study, and results in prolonged inflammatory response increasing morbidity. Even after influenza is cleared, *S. pneumoniae* bacterial clearance is affected. This study has highlighted some serotype dependent differences, suggesting different treatment programs would be beneficial for different serotypes. Evidence that it is worth further looking into co-infection of influenza with different serotypes of *S. pneumoniae* and other bacteria of interest (Sharma-Chawla et al., 2016). Impaired neutrophils have been shown to correlate with secondary bacterial infection

in Chinchillas, due to the importance of phagocytosis during innate immunity (Abramson et al., 1981). Influenza infection is known to result in the production of IFN $\gamma$ ; pulmonary IFN $\gamma$  pro-inflammatory cytokines are produced by natural killer (NK) cells as part of innate immunity and by CD4 and CD8 NK T cells as part of adaptive immunity (Schoenborn and Wilson, 2007). They increase macrophage activation during innate immunity (Scott et al., 2004) however, during T cell responses to viral infection they have been shown to inhibit bacterial clearance from the respiratory system by macrophages. It is thought that as they assist in the induction of specific anti-influenza adaptive immunity they down regulate innate immunity. The resulting suppression of phagocytosis paves the way for successful bacterial infection (Sun and Metzger, 2008). Additionally, Type I IFNs inhibit interleukin 23 (IL-23) dependent induction of T helper cell 17 (Th17) immunity, and therefore there is a reduction in the levels of CD4+ T cells and gamma delta T cells and hence a reduction in the production of IL-17 and IL-22, preventing the clearance of bacteria (Shahangian et al., 2009; Kudva et al., 2011; Mulcahy and McLoughlin, 2016). Robinson et al. (2013) have also shown that influenza A infection significantly decreased IL-1 $\beta$  production; IL-1 $\beta$  has been shown to play a role in Th17 polarization, therefore further hindering this pathway of immunity (Robinson et al., 2013). During co/secondary *S. pneumoniae* infection, type I IFNs have been shown to inhibit the production of specific chemokines (KC/CXCL1 and Mip2/CXCL2) resulting in an attenuated neutrophil response (Shahangian et al., 2009). Viral and bacterial co-infection of monocyte derived macrophages synergistically induces a pro-inflammatory response related to the type-I IFN and JAK-STAT signaling pathways (Hoffmann et al., 2016). Inflammation causes tissue damage, revealing more attachment sites for increased/developed bacterial infection. Co-infection also results in a synergistic increase in type II IFN (IFN $\gamma$ ) when compared to individual infection of influenza or *S. pneumoniae*, CXCL10 (aka IFN $\gamma$ -induced protein 10/IP-10) is secreted in response. IP-10 attracts various immune cells including activated T cells, monocytes and macrophages, therefore causing inflammation (Dufour et al., 2002; Hoffmann et al., 2016). Patients suffering severe pneumonia show significantly higher levels of IP-10 than those with minor cases of pneumonia. IP-10 is increased during *H. influenzae* and *S. aureus* co-infection as well, and like with *S. pneumoniae*, correlates with/highlights pneumonia etiology (Hoffmann et al., 2016). In addition, when *S. pneumoniae* successively co-infects with influenza it leads to severe clinical complications; partly due to an increase in apoptosis of dendritic cells, which therefore reduces T cell priming impairing the development of adaptive immunity. Influenza and *S. pneumoniae* infections can also lead to synergistic and non-synergistic dysregulation of cytokine responses (Wu et al., 2011).

As previously described influenza infection results in a reduction in the production of IL-17 and 22. IL-17 is important in the clearance of *S. aureus* by neutrophils (Archer et al., 2013). IL-22 is involved in controlling the production of antimicrobial peptides as well as staphylococcal ligand expression (Robinson et al., 2014; Mulcahy et al., 2016). In addition to



this, influenza positively affects the colonization of *S. aureus* by causing increased type III- IFN expression, which alters the IL-22 responses impairing host expression of antimicrobial peptides (Mulcahy and McLoughlin, 2016). Distress signals, such as ATP and norepinephrine, produced by damaged influenza also have several effects on *S. aureus*; namely the instigation of biofilm dispersal which helps the spread of *S. aureus* into the lungs, assisting in the development of pneumonia and invasive disease (Mulcahy and McLoughlin, 2016).

Of course viral infection doesn't just benefit bacteria; several mechanisms of synergism between viruses and bacteria have been suggested. A number of studies have documented an increase in viral load as viral clearance is reduced during bacterial co-infection. It is, however, unclear whether this is from bacterial and viral cooperation/interactions or simply from bacteria burdening the host immune system resulting in the reduction of viral eradication. Therefore further research is required to develop our understanding of the interaction between bacteria and viruses within co-infection (Peltola and McCullers, 2004; Iverson et al., 2011; Smith and McCullers, 2014).

## CONCLUSION

Viral infection aids bacterial infection in a number of ways, including unveiling/providing more sites for adhesion, impairing immune responses and causing cell and tissue destruction allowing for the spread of bacteria and development of invasive infection. Bacterial infection is then able to worsen clinical outcome and the severity of disease. Of course viral and bacterial co-infection can be mutually beneficial, further helping

viral infection, which is bad news for public health. Although antibiotics can help reduce the impact of co/secondary bacterial infection, we still need to better understand the interactions between viruses, bacteria and their host, and to fully understand all mechanisms of disease. Particularly in light of increased antibiotic resistance and the ability of microbes to adapt and evade vaccine induced immunity.

The aim of this review was to emphasize the historical and continuing threat of influenza and to highlight the risk of bacterial co/secondary infection. Vaccines and antibiotics are readily available, however, with antibiotic resistance at an all-time high, vaccination is becoming even more vital in the fight against influenza epidemics and pandemics and the bacterial co/secondary infections commonly associated. It is important to examine the strains and types of bacteria and viruses being spread amongst and transmitted throughout the general public (or continue to in the case of influenza) to inform clinical treatment and development, particularly in the setting of an influenza epidemic or pandemic. As the threat from influenza is ever changing, we need to ensure we know what strains are circulating, which could cause issue and how they interact with other potential pathogens. This preparation also entails monitoring the changing epidemiology of bacterial pathogens associated with secondary infection, such as capsule switching which help *S. pneumoniae* evade immunity (Pai et al., 2005a,b).

## AUTHOR CONTRIBUTIONS

DM, DC, and SC designed, planned and wrote the manuscript.

## REFERENCES

- Abdullahi, O., Nyiro, J., Lewa, P., Slack, M., and Scott, J. A. (2008). The descriptive epidemiology of *Streptococcus pneumoniae* and *Haemophilus influenzae* nasopharyngeal carriage in children and adults in Kilifi district, Kenya. *Pediatr. Infect. Dis. J.* 27, 59–64. doi: 10.1097/INF.0b013e31814da70c
- Abrahams, A., Hallows, N., and French, H. (1919). A further investigation into influenza-pneumococcal and influenza-streptococcal septicaemia: epidemic influenza "pneumonia" of highly fatal type and its relation to "purulent bronchitis" *Lancet* 193, 1–11. doi: 10.1016/S0140-6736(01)22115-1
- Abramson, J. S., Giebink, G. S., Mills, E. L., and Quie, P. G. (1981). Polymorphonuclear leukocyte dysfunction during influenza virus infection in chinchillas. *J. Infect. Dis.* 143, 836–845. doi: 10.1093/infdis/143.6.836
- Aebi, T., Weisser, M., Bucher, E., Hirsch, H. H., Marsch, S., and Siegemund, M. (2010). Co-infection of influenza B and *Streptococci* causing severe pneumonia and septic shock in healthy women. *BMC Infect. Dis.* 10:308. doi: 10.1186/1471-2334-10-308
- Archer, N. K., Harro, J. M., and Shirtliff, M. E. (2013). Clearance of *Staphylococcus aureus* nasal carriage is T cell dependent and mediated through interleukin-17A expression and neutrophil influx. *Infect. Immun.* 81, 2070–2075. doi: 10.1128/IAI.00084-13
- Baum, L. G., and Paulson, J. C. (1991). The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity. *Virology* 180, 10–15. doi: 10.1016/0042-6822(91)90003-T
- Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitz, E., Collins, M., et al. (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* 2:e31. doi: 10.1371/journal.pgen.0020031
- Berendt, R. F., Long, G. G., and Walker, J. S. (1975). Influenza alone and in sequence with pneumonia due to *Streptococcus pneumoniae* in the squirrel monkey. *J. Infect. Dis.* 132, 689–693. doi: 10.1093/infdis/132.6.689
- Berg, R. A., and Bartley, D. L. (1987). Pneumonia associated with *Branhamella catarrhalis* in infants. *Pediatr. Infect. Dis. J.* 6, 569–573. doi: 10.1097/00006454-198706000-00017
- Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., and Finelli, L. (2014). Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect. Dis.* 14:480. doi: 10.1186/1471-2334-14-480
- Blanton, W. B., and Irons, E. E. (1918). A recent epidemic of acute respiratory infection at camp custer, mich: preliminary laboratory report. *J. Am. Med. Assoc.* 71, 1988–1991. doi: 10.1001/jama.1918.26020500014006d
- Bluestone, C. D. (1986). Otitis media and sinusitis in children. Role of *Branhamella catarrhalis*. *Drugs* 31(Suppl. 3), 132–141.
- Brem, W. V., Bolling, G. E., and Casper, E. J. (1918). Pandemic "influenza" and secondary pneumonia at camp fremont, Calif. *J. Am. Med. Assoc.* 71, 2138–2144. doi: 10.1001/jama.1918.26020520007010b
- Bridy-Pappas, A. E., Margolis, M. B., Center, K. J., and Isaacman, D. J. (2005). *Streptococcus pneumoniae*: description of the pathogen, disease epidemiology, treatment, and prevention. *Pharmacotherapy* 25, 1193–1212. doi: 10.1592/phco.2005.25.9.1193
- Brouwer, M. C., Tunkel, A. R., and van de Beek, D. (2010). Epidemiology, diagnosis, and antimicrobial treatment of acute bacterial meningitis. *Clin. Microbiol. Rev.* 23, 467–492. doi: 10.1128/CMR.00070-09
- Brundage, J. F. (2006). Interactions between influenza and bacterial respiratory pathogens: implications for pandemic preparedness. *Lancet Infect. Dis.* 6, 303–312. doi: 10.1016/S1473-3099(06)70466-2

- Brundage, J. F., and Shanks, G. D. (2008). Deaths from bacterial pneumonia during 1918-19 influenza pandemic. *Emerg. Infect. Dis.* 14, 1193–1199. doi: 10.3201/eid1408.071313
- Bucior, I., Pielage, J. F., and Engel, J. N. (2012). *Pseudomonas aeruginosa* pili and flagella mediate distinct binding and signaling events at the apical and basolateral surface of airway epithelium. *PLoS Pathog.* 8:e1002616. doi: 10.1371/journal.ppat.1002616
- Burnet, F. M., and Clark, E. (1942). *Influenza: A Survey of the Last 50 Years in the Light of Modern Work on the Virus of Epidemic Influenza*. London: Macmillan and co Ltd.
- Calix, J. J., and Nahm, M. H. (2010). A new pneumococcal serotype, 11E, has a variably inactivated wjE gene. *J. Infect. Dis.* 202, 29–38. doi: 10.1086/653123
- Calix, J. J., Porambo, R. J., Brady, A. M., Larson, T. R., Yother, J., Abeygunwardana, C., et al. (2012). Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. *J. Biol. Chem.* 287, 27885–27894. doi: 10.1074/jbc.M112.380451
- Catlin, B. W. (1990). *Branhamella catarrhalis*: an organism gaining respect as a pathogen. *Clin. Microbiol. Rev.* 3, 293–320. doi: 10.1128/CMR.3.4.293
- Cauley, L. S., and Vella, A. T. (2015). Why is coinfection with influenza virus and bacteria so difficult to control? *Discov. Med.* 19, 33–40.
- Centers for Disease Control and Prevention (2009). Bacterial co-infection in lung tissue specimens from fatal cases of 2009 pandemic influenza A (H1N1) - United States, May–August 2009. *Morb. Mortal. Wkly. Rep.* 58, 1071–1074.
- Chen, R., and Holmes, E. C. (2006). Avian influenza virus exhibits rapid evolutionary dynamics. *Mol. Biol. Evol.* 23, 2336–2341. doi: 10.1093/molbev/msl102
- Chertow, D. S., and Memoli, M. J. (2013). Bacterial coinfection in influenza: a grand rounds review. *JAMA* 309, 275–282. doi: 10.1001/jama.2012.194139
- Chien, Y. W., Klugman, K. P., and Morens, D. M. (2009). Bacterial pathogens and death during the 1918 influenza pandemic. *N. Engl. J. Med.* 361, 2582–2583. doi: 10.1056/NEJMc0908216
- Chin, C. L., Manzel, L. J., Lehman, E. E., Humlicek, A. L., Shi, L., Starner, T. D., et al. (2005). *Haemophilus influenzae* from patients with chronic obstructive pulmonary disease exacerbation induce more inflammation than colonizers. *Am. J. Respir. Crit. Care Med.* 172, 85–91. doi: 10.1164/rccm.200412-1687OC
- Cockburn, W. C., Delon, P. J., and Ferreira, W. (1969). Origin and progress of the 1968-69 Hong Kong influenza epidemic. *Bull. World Health Organ.* 41, 345–348.
- Collazos, J., de Miguel, J., and Ayarza, R. (1992). *Moraxella catarrhalis* bacteremic pneumonia in adults: two cases and review of the literature. *Eur. J. Clin. Microbiol. Infect. Dis.* 11, 237–240. doi: 10.1007/BF02098086
- Conenello, G. M., and Palese, P. (2007). Influenza A virus PB1-F2: a small protein with a big punch. *Cell Host Microbe* 2, 207–209. doi: 10.1016/j.chom.2007.09.010
- Couch, R. B., Kasel, J. A., Gerin, J. L., Schulman, J. L., and Kilbourne, E. D. (1974). Induction of partial immunity to influenza by a neuraminidase-specific vaccine. *J. Infect. Dis.* 129, 411–420. doi: 10.1093/infdis/129.4.411
- Cox, N. J., and Subbarao, K. (2000). Global epidemiology of influenza: past and present. *Annu. Rev. Med.* 51, 407–421. doi: 10.1146/annurev.med.51.1.407
- Cundell, D. R., Gerard, N. P., Gerard, C., Idanpaan-Heikkilä, I., and Tuomanen, E. I. (1995). *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature* 377, 435–438. doi: 10.1038/377435a0
- Cunha, B. A., Pherez, F. M., and Schoch, P. (2009). Diagnostic importance of relative lymphopenia as a marker of swine influenza (H1N1) in adults. *Clin. Infect. Dis.* 49, 1454–1456. doi: 10.1086/644496
- Daszak, P. (2012). Anatomy of a pandemic. *Lancet* 380, 1883–1884. doi: 10.1016/S0140-6736(12)61887-X
- Davison, V. E., and Sanford, B. A. (1981). Adherence of *Staphylococcus aureus* to influenza A virus-infected madin-darby canine kidney cell cultures. *Infect. Immun.* 32, 118–126.
- de Jong, M. D., Simmons, C. P., Thanh, T. T., Hien, V. M., Smith, G. J., Chau, T. N., et al. (2006). Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat. Med.* 12, 1203–1207. doi: 10.1038/nm1477
- Dela Cruz, C. S., and Wunderink, R. G. (2017). *Respiratory Viral and Atypical Pneumonia in Adults*. Amsterdam: Elsevier.
- DeLeo, F. R., and Musser, J. M. (2010). Axis of coinfection evil. *J. Infect. Dis.* 201, 488–490. doi: 10.1086/650304
- Dennis Shanks, G., Mackenzie, A., Waller, M., and Brundage, J. F. (2012). Relationship between “purulent bronchitis” in military populations in Europe prior to 1918 and the 1918-1919 influenza pandemic. *Influenza Other Respir. Viruses* 6, 235–239. doi: 10.1111/j.1750-2659.2011.00309.x
- Dhanoa, A., Fang, N. C., Hassan, S. S., Kaniappan, P., and Rajasekaram, G. (2011). Epidemiology and clinical characteristics of hospitalized patients with pandemic influenza A (H1N1) 2009 infections: the effects of bacterial coinfection. *Viol. J.* 8:501. doi: 10.1186/1743-422X-8-501
- Dormitzer, P. R., Galli, G., Castellino, F., Golding, H., Khurana, S., Del Giudice, G., et al. (2011). Influenza vaccine immunology. *Immunol. Rev.* 239, 167–177. doi: 10.1111/j.1600-065X.2010.00974.x
- Dufour, J. H., Dziejman, M., Liu, M. T., Leung, J. H., Lane, T. E., and Luster, A. D. (2002). IFN-gamma-inducible protein 10 (IP-10; CXCL10)-deficient mice reveal a role for IP-10 in effector T cell generation and trafficking. *J. Immunol.* 168, 3195–3204. doi: 10.4049/jimmunol.168.7.3195
- Dunn, F. L. (1958). Pandemic influenza in 1957; review of international spread of new Asian strain. *J. Am. Med. Assoc.* 166, 1140–1148. doi: 10.1001/jama.1958.02990100028006
- Dupont, D., Mahjoub-Messai, F., Francois, M., Doit, C., Mariani-Kurkdjian, P., Bidet, P., et al. (2010). Evolving microbiology of complicated acute otitis media before and after introduction of the pneumococcal conjugate vaccine in France. *Diagn. Microbiol. Infect. Dis.* 68, 89–92. doi: 10.1016/j.diagmicrobio.2010.04.012
- Eldin, C., Melenotte, C., Mediannikov, O., Ghigo, E., Million, M., Edouard, S., et al. (2017). From Q fever to *Coxiella burnetii* infection: a paradigm change. *Clin. Microbiol. Rev.* 30, 115–190. doi: 10.1128/CMR.00045-16
- Eom, S.-H., Lee, D.-S., Jung, Y.-J., Park, J.-H., Choi, J.-I., Yim, M.-J., et al. (2014). The mechanism of antibacterial activity of phlorofucofuroeckol-A against methicillin-resistant *Staphylococcus aureus*. *Appl. Microbiol. Biotechnol.* 98, 9795–9804. doi: 10.1007/s00253-014-6041-8
- Faden, H., Harabuchi, Y., and Hong, J. J. (1994). Epidemiology of *Moraxella catarrhalis* in children during the first 2 years of life: relationship to otitis media. *J. Infect. Dis.* 169, 1312–1317. doi: 10.1093/infdis/169.6.1312
- Fairstein, V., Musher, D. M., and Cate, T. R. (1980). Bacterial adherence to pharyngeal cells during viral infection. *J. Infect. Dis.* 141, 172–176. doi: 10.1093/infdis/141.2.172
- Feder, H. M. Jr., and Garibaldi, R. A. (1984). The significance of nongonococcal, nonmeningococcal *Neisseria* isolates from blood cultures. *Rev. Infect. Dis.* 6, 181–188. doi: 10.1093/clinids/6.2.181
- Fedson, D. S., Wajda, A., Nicol, J. P., Hammond, G. W., Kaiser, D. L., and Roos, L. L. (1993). Clinical effectiveness of influenza vaccination in Manitoba. *JAMA* 270, 1956–1961. doi: 10.1001/jama.1993.03510160074032
- Finelli, L., Fiore, A., Dhara, R., Brammer, L., Shay, D. K., Kamimoto, L., et al. (2008). Influenza-associated pediatric mortality in the United States: increase of *Staphylococcus aureus* coinfection. *Pediatrics* 122, 805–811. doi: 10.1542/peds.2008-1336
- Fishovitz, J., Rojas-Altuve, A., Otero, L. H., Dawley, M., Carrasco-López, C., Chang, M., et al. (2014). Disruption of allosteric response as an unprecedented mechanism of resistance to antibiotics. *J. Am. Chem. Soc.* 136, 9814–9817. doi: 10.1021/ja5030657
- Fukumi, H. (1959). Summary report on the Asian influenza epidemic in Japan, 1957. *Bull. World Health Organ.* 20, 187–198.
- Gatherer, D. (2009). The 2009 H1N1 influenza outbreak in its historical context. *J. Clin. Virol.* 45, 174–178. doi: 10.1016/j.jcv.2009.06.004
- Gill, J. R., Sheng, Z. M., Ely, S. F., Guinee, D. G., Beasley, M. B., Suh, J., et al. (2010). Pulmonary pathologic findings of fatal 2009 pandemic influenza A/H1N1 viral infections. *Arch. Pathol. Lab. Med.* 134, 235–243. doi: 10.1043/1543-2165-134.2.235
- Glezen, W. P. (1996). Emerging infections: pandemic influenza. *Epidemiol. Rev.* 18, 64–76. doi: 10.1093/oxfordjournals.epirev.a017917
- Grijalva, C. G., Griffin, M. R., Edwards, K. M., Williams, J. V., Gil, A. I., Verastegui, H., et al. (2014). The role of influenza and parainfluenza infections in nasopharyngeal pneumococcal acquisition among young children. *Clin. Infect. Dis.* 58, 1369–1376. doi: 10.1093/cid/ciu148

- Gupta, R. K., George, R., and Nguyen-Van-Tam, J. S. (2008). Bacterial pneumonia and pandemic influenza planning. *Emerg. Infect. Dis.* 14, 1187–1192. doi: 10.3201/eid1408.070751
- Hager, H., Verghese, A., Alvarez, S., and Berk, S. L. (1987). Branhamella catarrhalis respiratory infections. *Rev. Infect. Dis.* 9, 1140–1149. doi: 10.1093/clinids/9.6.1140
- Hakansson, A., Kidd, A., Wadell, G., Sabharwal, H., and Svanborg, C. (1994). Adenovirus infection enhances in vitro adherence of *Streptococcus pneumoniae*. *Infect. Immun.* 62, 2707–2714.
- Hall, J. N., Stone, M. C., and Simpson, J. C. (1918). The epidemic of pneumonia following influenza at Camp Logan, Texas: preliminary report. *J. Am. Med. Assoc.* 71, 1986–1987. doi: 10.1001/jama.1918.26020500012006c
- Hament, J.-M., Kimpen, J. L. L., Fleer, A., and Wolfs, T. F. W. (1999). Respiratory viral infection predisposing for bacterial disease: a concise review. *FEMS Immunol. Med. Microbiol.* 26, 189–195. doi: 10.1111/j.1574-695X.1999.tb01389.x
- Hampson, A. W., and Mackenzie, J. S. (2006). The influenza viruses. *Med. J. Aust.* 185, S39–S43.
- Hers, J. F., Masurel, N., and Mulder, J. (1958). Bacteriology and histopathology of the respiratory tract and lungs in fatal Asian influenza. *Lancet* 2, 1141–1143. doi: 10.1016/S0140-6736(58)92404-8
- Hilleman, M. R. (2002). Realities and enigmas of human viral influenza: pathogenesis, epidemiology and control. *Vaccine* 20, 3068–3087. doi: 10.1016/S0264-410X(02)00254-2
- Hoffmann, J., Machado, D., Terrier, O., Pouzol, S., Messaoudi, M., Basualdo, W., et al. (2016). Viral and bacterial co-infection in severe pneumonia triggers innate immune responses and specifically enhances IP-10: a translational study. *Sci. Rep.* 6:38532. doi: 10.1038/srep38532
- Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. *Science* 215, 1577–1585. doi: 10.1126/science.7041255
- Hooper, L. V., Littman, D. R., and Macpherson, A. J. (2012). Interactions between the microbiota and the immune system. *Science* 336, 1268–1273. doi: 10.1126/science.1223490
- Houser, K., and Subbarao, K. (2015). Influenza vaccines: challenges and solutions. *Cell Host Microbe* 17, 295–300. doi: 10.1016/j.chom.2015.02.012
- Huang, S. S., Johnson, K. M., Ray, G. T., Wroe, P., Lieu, T. A., Moore, M. R., et al. (2011). Healthcare utilization and cost of pneumococcal disease in the United States. *Vaccine* 29, 3398–3412. doi: 10.1016/j.vaccine.2011.02.088
- Ioannidis, J. P., Worthington, M., Griffiths, J. K., and Snyderman, D. R. (1995). Spectrum and significance of bacteremia due to *Moraxella catarrhalis*. *Clin. Infect. Dis.* 21, 390–397. doi: 10.1093/clinids/21.2.390
- Iverson, A. R., Boyd, K. L., McAuley, J. L., Plano, L. R., Hart, M. E., and McCullers, J. A. (2011). Influenza virus primes mice for pneumonia from *Staphylococcus aureus*. *J. Infect. Dis.* 203, 880–888. doi: 10.1093/infdis/jiq113
- Jackson, C., Mangtani, P., Hawker, J., Olowokure, B., and Vynnycky, E. (2014). The effects of school closures on influenza outbreaks and pandemics: systematic review of simulation studies. *PLoS ONE* 9:e97297. doi: 10.1371/journal.pone.0097297
- Jacoby, P., Watson, K., Bowman, J., Taylor, A., Riley, T. V., Smith, D. W., et al. (2007). Modelling the co-occurrence of *Streptococcus pneumoniae* with other bacterial and viral pathogens in the upper respiratory tract. *Vaccine* 25, 2458–2464. doi: 10.1016/j.vaccine.2006.09.020
- Jakeman, K. J., Tisdale, M., Russell, S., Leone, A., and Sweet, C. (1994). Efficacy of 2'-deoxy-2'-fluororibosides against influenza A and B viruses in ferrets. *Antimicrob. Agents Chemother.* 38, 1864–1867. doi: 10.1128/AAC.38.8.1864
- Jin, P., Kong, F., Xiao, M., Oftadeh, S., Zhou, F., Liu, C., et al. (2009). First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. *J. Infect. Dis.* 200, 1375–1380. doi: 10.1086/606118
- Johnson, B. (2016). *The Spanish Flu Pandemic of 1918*. Available at: <http://www.historic-uk.com/HistoryUK/HistoryofBritain/The-Spanish-Flu-pandemic-of-1918/>
- Johnson, M. A., Drew, W. L., and Roberts, M. (1981). Branhamella (*Neisseria*) catarrhalis—a lower respiratory tract pathogen? *J. Clin. Microbiol.* 13, 1066–1069.
- Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993. doi: 10.1038/nature06536
- Joseph, C., Togawa, Y., and Shindo, N. (2013). Bacterial and viral infections associated with influenza. *Influenza Other Respir. Viruses* 7(Suppl. 2), 105–113. doi: 10.1111/irv.12089
- Karesh, W. B., Dobson, A., Lloyd-Smith, J. O., Lubroth, J., Dixon, M. A., Bennett, M., et al. (2012). Ecology of zoonoses: natural and unnatural histories. *Lancet* 380, 1936–1945. doi: 10.1016/S0140-6736(12)61678-X
- Kash, J. C., and Taubenberger, J. K. (2015). The role of viral, host, and secondary bacterial factors in influenza pathogenesis. *Am. J. Pathol.* 185, 1528–1536. doi: 10.1016/j.ajpath.2014.08.030
- Kash, J. C., Tumpey, T. M., Prohl, S. C., Carter, V., Perwitasari, O., Thomas, M. J., et al. (2006). Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* 443, 578–581. doi: 10.1038/nature05181
- Kash, J. C., Walters, K. A., Davis, A. S., Sandouk, A., Schwartzman, L. M., Jagger, B. W., et al. (2011). Lethal synergism of 2009 pandemic H1N1 influenza virus and *Streptococcus pneumoniae* coinfection is associated with loss of murine lung repair responses. *MBio* 2:e00172-11. doi: 10.1128/mBio.00172-11
- Kash, J. C., Xiao, Y., Davis, A. S., Walters, K. A., Chertow, D. S., Easterbrook, J. D., et al. (2014). Treatment with the reactive oxygen species scavenger EUK-207 reduces lung damage and increases survival during 1918 influenza virus infection in mice. *Free Radic. Biol. Med.* 67, 235–247. doi: 10.1016/j.freeradbiomed.2013.10.014
- Kilpi, T., Herva, E., Kajjalainen, T., Syrjanen, R., and Takala, A. K. (2001). Bacteriology of acute otitis media in a cohort of Finnish children followed for the first two years of life. *Pediatr. Infect. Dis. J.* 20, 654–662. doi: 10.1097/00006454-200107000-00004
- Klein, E. Y., Monteforte, B., Gupta, A., Jiang, W., May, L., Hsieh, Y. H., et al. (2016). The frequency of influenza and bacterial coinfection: a systematic review and meta-analysis. *Influenza Other Respir. Viruses* 10, 394–403. doi: 10.1111/irv.12398
- Kleiner, E. S., Daniels, C. A., Polissone, R. P., and Snyderman, R. (1976). Effect of virus infection on the inflammatory response. Depression of macrophage accumulation in influenza-infected mice. *Am. J. Pathol.* 85, 373–382.
- Klimov, A., Simonsen, L., Fukuda, K., and Cox, N. (1999). Surveillance and impact of influenza in the United States. *Vaccine* 17(Suppl. 1), S42–S46. doi: 10.1016/S0264-410X(99)00104-8
- Kobasa, D., Jones, S. M., Shinya, K., Kash, J. C., Copps, J., Ebihara, H., et al. (2007). Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature* 445, 319–323. doi: 10.1038/nature05495
- Kollef, M. H., Shorr, A., Tabak, Y. P., Gupta, V., Liu, L. Z., and Johannes, R. S. (2005). Epidemiology and outcomes of health-care-associated pneumonia: results from a large US database of culture-positive pneumonia. *Chest* 128, 3854–3862. doi: 10.1378/chest.128.6.3854
- Kudva, A., Scheller, E. V., Robinson, K. M., Crowe, C. R., Choi, S. M., Slight, S. R., et al. (2011). Influenza A inhibits Th17-mediated host defense against bacterial pneumonia in mice. *J. Immunol.* 186, 1666–1674. doi: 10.4049/jimmunol.1002194
- Lamkanfi, M., and Dixit, V. M. (2010). Manipulation of Host Cell Death Pathways during Microbial Infections. *Cell Host Microbe* 8, 44–54. doi: 10.1016/j.chom.2010.06.007
- Langereis, J. D., and de Jonge, M. I. (2015). Invasive disease caused by nontypeable *Haemophilus influenzae*. *Emerg. Infect. Dis.* 21, 1711–1718. doi: 10.3201/eid2110.150004
- Lee, E. H., Wu, C., Lee, E. U., Stoute, A., Hanson, H., Cook, H. A., et al. (2010). Fatalities associated with the 2009 H1N1 influenza A virus in New York city. *Clin. Infect. Dis.* 50, 1498–1504. doi: 10.1086/652446
- Lee, L. N., Dias, P., Han, D., Yoon, S., Shea, A., Zakharov, V., et al. (2010). A mouse model of lethal synergism between influenza virus and *Haemophilus influenzae*. *Am. J. Pathol.* 176, 800–811. doi: 10.2353/ajpath.2010.090596
- Lee, M. H., Arrecubieta, C., Martin, F. J., Prince, A., Borczuk, A. C., and Lowy, F. D. (2010). A postinfluenza model of *Staphylococcus aureus* pneumonia. *J. Infect. Dis.* 201, 508–515. doi: 10.1086/650204
- Lee, P. H., Bird, N., MacKenzie-Kludas, C., Mansell, A., Kedzierska, K., Brown, L., et al. (2016). Induction of memory cytotoxic T cells to influenza A virus and subsequent viral clearance is not modulated by PB1-F2-dependent inflammasome activation. *Immunol. Cell Biol.* 94, 439–446. doi: 10.1038/icb.2015.115



- Lee, S. E., Eick, A., Bloom, M. S., and Brundage, J. F. (2008). Influenza immunization and subsequent diagnoses of group A streptococcus-illnesses among U.S. Army trainees, 2002–2006. *Vaccine* 26, 3383–3386. doi: 10.1016/j.vaccine.2008.04.041
- Li, Q., Sun, X., Li, Z., Liu, Y., Vavricka, C. J., Qi, J., et al. (2012). Structural and functional characterization of neuraminidase-like molecule N10 derived from bat influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18897–18902. doi: 10.1073/pnas.1211037109
- Lindsay, M. I. Jr., Herrmann, E. C. Jr., Morrow, G. W. Jr., and Brown, A. L. Jr. (1970). Hong kong influenza: clinical, microbiologic, and pathologic features in 127 cases. *JAMA* 214, 1825–1832. doi: 10.1001/jama.1970.03180100019004
- Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380, 2095–2128. doi: 10.1016/S0140-6736(12)61728-0
- Lucas, S. (2010). Predictive clinicopathological features derived from systematic autopsy examination of patients who died with A/H1N1 influenza infection in the UK 2009–10 pandemic. *Health Technol. Assess.* 14, 83–114. doi: 10.3310/hta14550-02
- Madhi, S. A., Adrian, P., Kuwanda, L., Cutland, C., Albrich, W. C., and Klugman, K. P. (2007). Long-term effect of pneumococcal conjugate vaccine on nasopharyngeal colonization by *Streptococcus pneumoniae* and associated interactions with *Staphylococcus aureus* and *Haemophilus influenzae* colonization in HIV-Infected and HIV-uninfected children. *J. Infect. Dis.* 196, 1662–1666. doi: 10.1086/522164
- Marchant, C. D. (1990). Spectrum of disease due to *Branhamella catarrhalis* in children with particular reference to acute otitis media. *Am. J. Med.* 88, 15S–19S. doi: 10.1016/0002-9343(90)90255-c
- Matsuzaki, Y., Sugawara, K., Mizuta, K., Tsuchiya, E., Muraki, Y., Hongo, S., et al. (2002). Antigenic and genetic characterization of influenza C viruses which caused two outbreaks in Yamagata City, Japan, in 1996 and 1998. *J. Clin. Microbiol.* 40, 422–429. doi: 10.1128/JCM.40.2.422-429.2002
- McCullers, J. A., and Bartmess, K. C. (2003). Role of neuraminidase in lethal synergism between influenza virus and *Streptococcus pneumoniae*. *J. Infect. Dis.* 187, 1000–1009. doi: 10.1086/368163
- McCullers, J. A., Iverson, A. R., McKeon, R., and Murray, P. J. (2008). The platelet activating factor receptor is not required for exacerbation of bacterial pneumonia following influenza. *Scand. J. Infect. Dis.* 40, 11–17. doi: 10.1080/00365540701477568
- McCullers, J. A., McAuley, J. L., Browall, S., Iverson, A. R., Boyd, K. L., and Henriques Normark, B. (2010). Influenza enhances susceptibility to natural acquisition of and disease due to *Streptococcus pneumoniae* in ferrets. *J. Infect. Dis.* 202, 1287–1295. doi: 10.1086/656333
- McCullers, J. A., and Reh, J. E. (2002). Lethal synergism between influenza virus and *Streptococcus pneumoniae*: characterization of a mouse model and the role of platelet-activating factor receptor. *J. Infect. Dis.* 186, 341–350. doi: 10.1086/341462
- McLeod, D. T., Ahmad, F., Power, J. T., Calder, M. A., and Seaton, A. (1983). Bronchopulmonary infection due to *Branhamella catarrhalis*. *Br. Med. J. (Clin. Res. Ed)* 287, 1446–1447. doi: 10.1136/bmj.287.6403.1446
- McNeely, D. J., Kitchens, C. S., and Kluge, R. M. (1976). Fatal *Neisseria (Branhamella) catarrhalis* pneumonia in an immunodeficient host. *Am. Rev. Respir. Dis.* 114, 399–402.
- Metersky, M. L., Masterton, R. G., Lode, H., File, T. M. Jr., and Babinchak, T. (2012). Epidemiology, microbiology, and treatment considerations for bacterial pneumonia complicating influenza. *Int. J. Infect. Dis.* 16, e321–31. doi: 10.1016/j.ijid.2012.01.003
- Michael, M. Jr. (1942). *Staphylococcus aureus* pneumonia: with special reference to its occurrence as a complication of influenza. *J. Am. Med. Assoc.* 118, 869–874. doi: 10.1001/jama.1942.02830110011004
- Michaelis, M., Doerr, H. W., and Cinatl, J. Jr. (2009). Novel swine-origin influenza A virus in humans: another pandemic knocking at the door. *Med. Microbiol. Immunol.* 198, 175–183. doi: 10.1007/s00430-009-0118-5
- Michaels, R. H., Myerowitz, R. L., and Klaw, R. (1977). Potentiation of experimental meningitis due to *Haemophilus influenzae* by influenza A virus. *J. Infect. Dis.* 135, 641–645. doi: 10.1093/infdis/135.4.641
- Morens, D. M., Taubenberger, J. K., and Fauci, A. S. (2008). Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J. Infect. Dis.* 198, 962–970. doi: 10.1086/591708
- Morse, S. S., Mazet, J. A., Woolhouse, M., Parrish, C. R., Carroll, D., Karesh, W. B., et al. (2012). Prediction and prevention of the next pandemic zoonosis. *Lancet* 380, 1956–1965. doi: 10.1016/S0140-6736(12)61684-5
- Mulcahy, M. E., Leech, J. M., Renaud, J. C., Mills, K. H., and McLoughlin, R. M. (2016). Interleukin-22 regulates antimicrobial peptide expression and keratinocyte differentiation to control *Staphylococcus aureus* colonization of the nasal mucosa. *Mucosal Immunol.* 9, 1429–1441. doi: 10.1038/mi.2016.24
- Mulcahy, M. E., and McLoughlin, R. M. (2016). *Staphylococcus aureus* and Influenza A Virus: partners in coinfection. *MBio* 7:e2068-16. doi: 10.1128/mBio.02068-16
- Murphy, F. A. (1998). Emerging zoonoses. *Emerg. Infect. Dis.* 4, 429–435. doi: 10.3201/eid0403.980324
- Murphy, T. F. (2003). Respiratory infections caused by non-typeable *Haemophilus influenzae*. *Curr. Opin. Infect. Dis.* 16, 129–134. doi: 10.1097/00001432-200304000-00009
- Murphy, T. F., Brauer, A. L., Grant, B. J., and Sethi, S. (2005). *Moraxella catarrhalis* in chronic obstructive pulmonary disease: burden of disease and immune response. *Am. J. Respir. Crit. Care Med.* 172, 195–199. doi: 10.1164/rccm.200412-1747OC
- Murray, R. J., Robinson, J. O., White, J. N., Hughes, F., Coombs, G. W., Pearson, J. C., et al. (2010). Community-acquired pneumonia due to pandemic A(H1N1)2009 influenza virus and methicillin resistant *Staphylococcus aureus* co-infection. *PLoS ONE* 5:e8705. doi: 10.1371/journal.pone.0008705
- Nakajima, K., Desselberger, U., and Palese, P. (1978). Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* 274, 334–339. doi: 10.1038/274334a0
- Nayak, D. P., Balogun, R. A., Yamada, H., Zhou, Z. H., and Barman, S. (2009). Influenza virus morphogenesis and budding. *Virus Res.* 143, 147–161. doi: 10.1016/j.virusres.2009.05.010
- Newing, W. J., and Christie, R. (1947). Meningitis; isolation of an organism resembling *Neisseria catarrhalis* from cerebro-spinal fluid; report of a case. *Med. J. Aust.* 1, 306.
- Nguyen, T., Kyle, U. G., Jaimon, N., Tchamitchi, M. H., Coss-Bu, J. A., Lam, F., et al. (2012). Coinfection with *Staphylococcus aureus* increases risk of severe coagulopathy in critically ill children with influenza A (H1N1) virus infection. *Crit. Care Med.* 40, 3246–3250. doi: 10.1097/CCM.0b013e318260c7f8
- Nichol, K. L., Wuorenma, J., and von Sternberg, T. (1998). Benefits of influenza vaccination for low-, intermediate-, and high-risk senior citizens. *Arch. Intern. Med.* 158, 1769–1776. doi: 10.1001/archinte.158.16.1769
- Okamoto, S., Kawabata, S., Nakagawa, I., Okuno, Y., Goto, T., Sano, K., et al. (2003). Influenza A virus-infected hosts boost an invasive type of *Streptococcus pyogenes* infection in mice. *J. Virol.* 77, 4104–4112. doi: 10.1128/JVI.77.7.4104-4112.2003
- Olson, D. R., Simonsen, L., Edelson, P. J., and Morse, S. S. (2005). Epidemiological evidence of an early wave of the 1918 influenza pandemic in New York City. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11059–11063. doi: 10.1073/pnas.0408290102
- Onofrio, J. M., Shulkin, A. N., Heidbrink, P. J., Toews, G. B., and Pierce, A. K. (1981). Pulmonary clearance and phagocytic cell response to normal pharyngeal flora. *Am. Rev. Respir. Dis.* 123, 222–225.
- Opie, E. L., Blake, F., Small, J. C., and Rivers, T. M. (1921). “The pneumonias and other infections of the respiratory tract accompanying influenza and measles. St. Louis: CV Mosby,” in *Epidemic Respiratory Disease*, eds E. L. Opie, F. G. Blake, J. C. Small, and T. M. Rivers (New York, NY: Forgotten Books), 107–281.
- Osterhaus, A. D., Rimmelzwaan, G. F., Martina, B. E., Bestebroer, T. M., and Fouchier, R. A. (2000). Influenza B virus in seals. *Science* 288, 1051–1053. doi: 10.1126/science.288.5468.1051
- Oxford, J. S., Sefton, A., Jackson, R., Innes, W., Daniels, R. S., and Johnson, N. P. (2002). World war I may have allowed the emergence of “Spanish” influenza. *Lancet Infect. Dis.* 2, 111–114. doi: 10.1016/S1473-3099(02)00185-8
- Pai, R., Gertz, R. E., Whitney, C. G., and Beall, B. (2005a). Clonal association between *Streptococcus pneumoniae* serotype 23A, circulating within the United States, and an internationally dispersed clone of serotype 23F. *J. Clin. Microbiol.* 43, 5440–5444.
- Pai, R., Moore, M. R., Pilishvili, T., Gertz, R. E., Whitney, C. G., Beall, B., et al. (2005b). Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. *J. Infect. Dis.* 192, 1988–1995.



- Palacios, G., Hornig, M., Cisterna, D., Savji, N., Bussetti, A. V., Kapoor, V., et al. (2009). *Streptococcus pneumoniae* coinfection is correlated with the severity of H1N1 pandemic influenza. *PLoS ONE* 4:e8540. doi: 10.1371/journal.pone.0008540
- Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C., and Nahm, M. H. (2007). Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* 45, 1225–1233. doi: 10.1128/JCM.02199-06
- Parrish, C. R., Murcia, P. R., and Holmes, E. C. (2015). Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans. *J. Virol.* 89, 2990–2994. doi: 10.1128/JVI.03146-14
- Payne, A. M. (1958). Some aspects of the epidemiology of the 1957 influenza pandemic. *Proc. R. Soc. Med.* 51, 1009–1015.
- Peckham, R. (2016). *Epidemics in Modern Asia*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781316026939
- Peiris, J. S., Poon, L. L., and Guan, Y. (2009). Emergence of a novel swine-origin influenza A virus (S-OIV) H1N1 virus in humans. *J. Clin. Virol.* 45, 169–173. doi: 10.1016/j.jcv.2009.06.006
- Peltola, V. T., and McCullers, J. A. (2004). Respiratory viruses predisposing to bacterial infections: role of neuraminidase. *Pediatr. Infect. Dis. J.* 23, S87–S97. doi: 10.1097/01.inf.0000108197.81270.35
- Peltola, V. T., Murti, K. G., and McCullers, J. A. (2005). Influenza virus neuraminidase contributes to secondary bacterial pneumonia. *J. Infect. Dis.* 192, 249–257. doi: 10.1086/430954
- Petersdorf, R. G., Fusco, J. J., Harter, D. H., and Albrink, W. S. (1959). Pulmonary infections complicating Asian influenza. *AMA Arch. Intern. Med.* 103, 262–272. doi: 10.1001/archinte.1959.00270020090010
- Pittet, L. F., and Posfay-Barbe, K. M. (2012). Pneumococcal vaccines for children: a global public health priority. *Clin. Microbiol. Infect.* 18(Suppl. 5), 25–36. doi: 10.1111/j.1469-0691.2012.03938.x
- Plotkowski, M. C., Puchelle, E., Beck, G., Jacquot, J., and Hannoun, C. (1986). Adherence of type I *Streptococcus pneumoniae* to tracheal epithelium of mice infected with influenza A/PR8 virus. *Am. Rev. Respir. Dis.* 134, 1040–1044. doi: 10.1164/arrd.1986.134.5.1040
- Potter, C. W. (2001). A history of influenza. *J. Appl. Microbiol.* 91, 572–579. doi: 10.1046/j.1365-2672.2001.01492.x
- Racaniello, V. (2009). *The A, B, and C of Influenza Virus*. Available at: <http://www.virology.ws/2009/09/22/the-a-b-and-c-of-influenza-virus/>
- Randolph, A. G., Vaughn, F., Sullivan, R., Robinson, L., Thompson, B. T., Yoon, G., et al. (2011). Critically ill children during the 2009–2010 influenza pandemic in the United States. *Pediatrics* 128:e1450–8. doi: 10.1542/peds.2011-0774
- Reid, A. H., Taubenberger, J. K., and Fanning, T. G. (2001). The 1918 Spanish influenza: integrating history and biology. *Microbes Infect.* 3, 81–87. doi: 10.1016/S1286-4579(00)01351-4
- Rice, T. W., Robinson, L., Uyeki, T. M., Vaughn, F. L., John, B. B., Miller, R. R., et al. (2012). Critical illness from 2009 pandemic influenza A virus and bacterial coinfection in the United States. *Crit. Care Med.* 40, 1487–1498. doi: 10.1097/CCM.0b013e3182416f23
- Rizzo, C., Caporali, M. G., and Rota, M. C. (2010). Pandemic influenza and pneumonia due to *Legionella pneumophila*: a frequently underestimated coinfection. *Clin. Infect. Dis.* 51:115. doi: 10.1086/653444
- Robertson, L., Caley, J. P., and Moore, J. (1958). Importance of *Staphylococcus aureus* in pneumonia in the 1957 epidemic of influenza A. *Lancet* 2, 233–236. doi: 10.1016/S0140-6736(58)90060-6
- Robinson, K. M., Choi, S. M., McHugh, K. J., Mandalapu, S., Enelow, R. I., Kolls, J. K., et al. (2013). Influenza A exacerbates *Staphylococcus aureus* pneumonia by attenuating IL-1 $\beta$  production in mice. *J. Immunol.* 191, 5153–5159. doi: 10.4049/jimmunol.1301237
- Robinson, K. M., Kolls, J. K., and Alcorn, J. F. (2015). The immunology of influenza virus-associated bacterial pneumonia. *Curr. Opin. Immunol.* 34, 59–67. doi: 10.1016/j.coi.2015.02.002
- Robinson, K. M., McHugh, K. J., Mandalapu, S., Clay, M. E., Lee, B., Scheller, E. V., et al. (2014). Influenza A virus exacerbates *Staphylococcus aureus* pneumonia in mice by attenuating antimicrobial peptide production. *J. Infect. Dis.* 209, 865–875. doi: 10.1093/infdis/jit527
- Rock, K. L., and Kono, H. (2008). The inflammatory response to cell death. *Annu. Rev. Pathol.* 3, 99–126. doi: 10.1146/annurev.pathmechdis.3.121806.151456
- Rosenstein, N. E., and Perkins, B. A. (2000). Update on *Haemophilus influenzae* serotype b and meningococcal vaccines. *Pediatr. Clin. North Am.* 47, 337–352. doi: 10.1016/S0031-3955(05)70210-8
- Rupp, S., Ambata, P., Narat, V., and Giles-Vernick, T. (2016). Beyond the cut hunter: a historical epidemiology of HIV beginnings in central africa. *Ecohealth* 13, 661–671. doi: 10.1007/s10393-016-1189-6
- Samji, T. (2009). Influenza A: understanding the viral life cycle. *Yale J. Biol. Med.* 82, 153–159.
- Sanford, B. A., and Ramsay, M. A. (1987). Bacterial adherence to the upper respiratory tract of ferrets infected with influenza A virus. *Proc. Soc. Exp. Biol. Med.* 185, 120–128. doi: 10.3181/00379727-185-42525
- Sanford, B. A., Shelokov, A., and Ramsay, M. A. (1978). Bacterial adherence to virus-infected cells: a cell culture model of bacterial superinfection. *J. Infect. Dis.* 137, 176–181. doi: 10.1093/infdis/137.2.176
- Schoenborn, J. R., and Wilson, C. B. (2007). Regulation of interferon-gamma during innate and adaptive immune responses. *Adv. Immunol.* 96, 41–101. doi: 10.1016/S0065-2776(07)96002-2
- Schwarzmann, S. W., Adler, J. L., Sullivan, R. J. Jr., and Marine, W. M. (1971). Bacterial pneumonia during the Hong Kong influenza epidemic of 1968–1969. *Arch. Intern. Med.* 127, 1037–1041. doi: 10.1001/archinte.1971.00310180053006
- Scott, M. J., Hoth, J. J., Stagner, M. K., Gardner, S. A., Peyton, J. C., and Cheadle, W. G. (2004). CD40-CD154 interactions between macrophages and natural killer cells during sepsis are critical for macrophage activation and are not interferon gamma dependent. *Clin. Exp. Immunol.* 137, 469–477. doi: 10.1111/j.1365-2249.2004.02547.x
- Seki, M., Yanagihara, K., Higashiyama, Y., Fukuda, Y., Kaneko, Y., Ohno, H., et al. (2004). Immunokinetics in severe pneumonia due to influenza virus and bacteria coinfection in mice. *Eur. Respir. J.* 24, 143–149. doi: 10.1183/09031936.04.00126103
- Selinger, D. S., Reed, W. P., and McLaren, L. C. (1981). Model for studying bacterial adherence to epithelial cells infected with viruses. *Infect. Immun.* 32, 941–944.
- Shahangian, A., Chow, E. K., Tian, X., Kang, J. R., Ghaffari, A., Liu, S. Y., et al. (2009). Type I IFNs mediate development of postinfluenza bacterial pneumonia in mice. *J. Clin. Invest.* 119, 1910–1920. doi: 10.1172/JCI35412
- Sharma-Chawla, N., Sender, V., Kershaw, O., Gruber, A. D., Volckmar, J., Henriques-Normark, B., et al. (2016). Influenza A virus infection predisposes hosts to secondary infection with different *Streptococcus pneumoniae* serotypes with similar outcome but serotype-specific manifestation. *Infect. Immun.* 84, 3445–3457. doi: 10.1128/IAI.00422-16
- Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., and Kawaoka, Y. (2006). Avian flu: influenza virus receptors in the human airway. *Nature* 440, 435–436. doi: 10.1038/440435a
- Shope, R. E. (1931). Swine Influenza: Iii. Filtration experiments and etiology. *J. Exp. Med.* 54, 373–385. doi: 10.1084/jem.54.3.373
- Shrestha, S., Foxman, B., Weinberger, D. M., Steiner, C., Viboud, C., and Rohani, P. (2013). Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Sci. Transl. Med.* 5:191ra184. doi: 10.1126/scitranslmed.3005982
- Siegel, S. J., Roche, A. M., and Weiser, J. N. (2014). Influenza promotes pneumococcal growth during coinfection by providing host sialylated substrates as a nutrient source. *Cell Host Microbe* 16, 55–67. doi: 10.1016/j.chom.2014.06.005
- Smith, A. M., and McCullers, J. A. (2014). Secondary bacterial infections in influenza virus infection pathogenesis. *Curr. Top. Microbiol. Immunol.* 385, 327–356. doi: 10.1007/82\_2014\_394
- Smith, F. B. (1995). The Russian influenza in the United Kingdom, 1889–1894. *Soc. Hist. Med.* 8, 55–73. doi: 10.1093/shm/8.1.55
- Spooner, L. H., Scott, J. M., and Heath, E. H. (1919). A bacteriologic study of the influenza epidemic at Camp Devens. *JAMA* 72, 155–159. doi: 10.1001/jama.1919.02610030001001
- Steinhauer, D. A. (1999). Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* 258, 1–20. doi: 10.1006/viro.1999.9716
- Sullivan, K. M., Monto, A. S., and Longini, I. M. Jr. (1993). Estimates of the US health impact of influenza. *Am. J. Public Health* 83, 1712–1716. doi: 10.2105/AJPH.83.12.1712
- Sun, K., and Metzger, D. W. (2008). Inhibition of pulmonary antibacterial defense by interferon-gamma during recovery from influenza infection. *Nat. Med.* 14, 558–564. doi: 10.1038/nm1765

- Takase, H., Nitanai, H., Yamamura, E., and Otani, T. (1999). Facilitated expansion of pneumococcal colonization from the nose to the lower respiratory tract in mice preinfected with influenza virus. *Microbiol. Immunol.* 43, 905–907. doi: 10.1111/j.1348-0421.1999.tb01226.x
- Tashiro, M., Ciborowski, P., Reinacher, M., Pulverer, G., Klenk, H. D., and Rott, R. (1987). Synergistic role of staphylococcal proteases in the induction of influenza virus pathogenicity. *Virology* 157, 421–430. doi: 10.1016/0042-6822(87)90284-4
- Taubenberger, J. K., and Morens, D. M. (2008). The pathology of influenza virus infections. *Annu. Rev. Pathol.* 3, 499–522. doi: 10.1146/annurev.pathmechdis.3.121806.154316
- Taubenberger, J. K., Reid, A. H., Janczewski, T. A., and Fanning, T. G. (2001). Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1829–1839. doi: 10.1098/rstb.2001.1020
- Tillett, H. E., Smith, J. W., and Gooch, C. D. (1983). Excess deaths attributable to influenza in England and Wales: age at death and certified cause. *Int. J. Epidemiol.* 12, 344–352. doi: 10.1093/ije/12.3.344
- Tong, S., Li, Y., Rivaller, P., Conrardy, C., Castillo, D. A., Chen, L. M., et al. (2012). A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4269–4274. doi: 10.1073/pnas.1116200109
- Tong, S. Y., Davis, J. S., Eichenberger, E., Holland, T. L., Fowler, V. G. Jr., et al. (2015). *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* 28, 603–661. doi: 10.1128/CMR.00134-14
- Topham, N. J., and Hewitt, E. W. (2009). Natural killer cell cytotoxicity: how do they pull the trigger? *Immunology* 128, 7–15. doi: 10.1111/j.1365-2567.2009.03123.x
- Trilla, A., Trilla, G., and Daer, C. (2008). The 1918 “Spanish Flu” in Spain. *Clin. Infect. Dis.* 47, 668–673. doi: 10.1086/590567
- Valleron, A. J., Cori, A., Valtat, S., Meurisse, S., Carrat, F., and Boelle, P. Y. (2010). Transmissibility and geographic spread of the 1889 influenza pandemic. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8778–8781. doi: 10.1073/pnas.1000886107
- van der Sluijs, K. F., van der Poll, T., Lutter, R., Juffermans, N. P., and Schultz, M. J. (2010). Bench-to-bedside review: bacterial pneumonia with influenza - pathogenesis and clinical implications. *Crit. Care* 14:219. doi: 10.1186/cc8893
- Verduin, C. M., Hol, C., Fleer, A., van Dijk, H., and van Belkum, A. (2002). *Moraxella catarrhalis*: from emerging to established pathogen. *Clin. Microbiol. Rev.* 15, 125–144. doi: 10.1128/CMR.15.1.125-144.2002
- Wadowsky, R. M., Mietzner, S. M., Skoner, D. P., Doyle, W. J., and Fireman, P. (1995). Effect of experimental influenza A virus infection on isolation of *Streptococcus pneumoniae* and other aerobic bacteria from the oropharynx of allergic and nonallergic adult subjects. *Infect. Immun.* 63, 1153–1157.
- Wang, T. T., and Palese, P. (2009). Unraveling the mystery of swine influenza virus. *Cell* 137, 983–985. doi: 10.1016/j.cell.2009.05.032
- Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56, 152–179.
- Weinberger, D. M., Simonsen, L., Jordan, R., Steiner, C., Miller, M., and Viboud, C. (2012). Impact of the 2009 influenza pandemic on pneumococcal pneumonia hospitalizations in the United States. *J. Infect. Dis.* 205, 458–465. doi: 10.1093/infdis/jir749
- Wenger, J. D. (1998). Epidemiology of Haemophilus influenzae type b disease and impact of Haemophilus influenzae type b conjugate vaccines in the United States and Canada. *Pediatr. Infect. Dis. J.* 17(9 Suppl.), S132–S136.
- Wertheim, H. F., Melles, D. C., Vos, M. C., van Leeuwen, W., van Belkum, A., Verbrugh, H. A., et al. (2005). The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect. Dis.* 5, 751–762. doi: 10.1016/S1473-3099(05)70295-4
- World Health Organization (2016). *Influenza (Seasonal) Factsheet*. Available at: <http://www.who.int/mediacentre/factsheets/fs211/en/>
- World Health Organization (2017). *Smallpox*. Available at: <http://web.archive.org/web/20070921235036/http://www.who.int/mediacentre/factsheets/smallpox/en/>
- Wu, S. W., de Lencastre, H., and Tomasz, A. (2005). Expression of high-level methicillin resistance in *Staphylococcus aureus* from the *Staphylococcus sciuri* mec A homologue: role of mutation(s) in the genetic background and in the coding region of mec A. *Microb. Drug Resist.* 11, 215–224. doi: 10.1089/mdr.2005.11.215
- Wu, Y., Mao, H., Ling, M. T., Chow, K. H., Ho, P. L., Tu, W., et al. (2011). Successive influenza virus infection and *Streptococcus pneumoniae* stimulation alter human dendritic cell function. *BMC Infect. Dis.* 11:201. doi: 10.1186/1471-2334-11-201
- Wu, Y., Wu, Y., Tefsen, B., Shi, Y., and Gao, G. F. (2014). Bat-derived influenza-like viruses H17N10 and H18N11. *Trends Microbiol.* 22, 183–191. doi: 10.1016/j.tim.2014.01.010
- Wyde, P. R., Six, H. R., Ambrose, M. W., and Throop, B. J. (1989). Influenza virus infection and bacterial clearance in young adult and aged mice. *J. Gerontol.* 44, B118–B124. doi: 10.1093/geronj/44.5.b118
- Yamashita, M., Krystal, M., Fitch, W. M., and Palese, P. (1988). Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. *Virology* 163, 112–122. doi: 10.1016/0042-6822(88)90238-3
- Yang, M., Gao, H., Chen, J., Xu, X., Tang, L., Yang, Y., et al. (2016a). Bacterial coinfection is associated with severity of avian influenza A (H7N9), and procalcitonin is a useful marker for early diagnosis. *Diagn. Microbiol. Infect. Dis.* 84, 165–169. doi: 10.1016/j.diagmicrobio.2015.10.018
- Yang, R., Cui, Y., and Bi, Y. (2016b). Perspectives on *Yersinia pestis*: a model for studying zoonotic pathogens. *Adv. Exp. Med. Biol.* 918, 377–391.
- Youzbashi, E., Marshall, M., Chaloupka, I., and Meier-Ewert, H. (1996). [Distribution of influenza C virus infection in dogs and pigs in Bavaria]. *Tierarztl. Prax.* 24, 337–342.
- Zambon, M. C. (2001). The pathogenesis of influenza in humans. *Rev. Med. Virol.* 11, 227–241. doi: 10.1002/rmv.319
- Zumla, A., Hui, D. S., Al-Tawfiq, J. A., Gautret, P., McCloskey, B., and Memish, Z. A. (2014). Emerging respiratory tract infections. *Lancet Infect. Dis.* 14, 910–911. doi: 10.1016/S1473-3099(14)70899-0

**Conflict of Interest Statement:** SC acts as principal investigator for clinical trials and other studies sponsored by the University Hospital Southampton NHS Foundation Trust/University of Southampton that are funded by vaccine manufacturers but receives no personal payments from them. SC has also participated in advisory boards for vaccine manufacturers but receives no personal payments for this work. SC has received financial assistance from vaccine manufacturers to attend conferences. All grants and honoraria are paid into accounts within the University of Southampton, or to independent charities. DC was employed for 18 months on a GSK funded research project in 2014/15.

The other author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Morris, Cleary and Clarke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Pivotal Role of DNA Repair in Infection Mediated-Inflammation and Cancer

Ayşe Z. Sahan<sup>1</sup>, Tapas K. Hazra<sup>2</sup> and Soumita Das<sup>1\*</sup>

<sup>1</sup> Department of Pathology, University of California, San Diego, San Diego, CA, United States, <sup>2</sup> Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX, United States

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Sultan Zainal Abidin University,  
Malaysia

### Reviewed by:

M. Gabriela Kramer,  
University of the Republic, Uruguay  
Arijit Dutta,  
Yale University, United States

### \*Correspondence:

Soumita Das  
sodas@ucsd.edu

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 03 November 2017

**Accepted:** 21 March 2018

**Published:** 11 April 2018

### Citation:

Sahan AZ, Hazra TK and Das S (2018)  
The Pivotal Role of DNA Repair in  
Infection Mediated-Inflammation and  
Cancer. *Front. Microbiol.* 9:663.  
doi: 10.3389/fmicb.2018.00663

Pathogenic and commensal microbes induce various levels of inflammation and metabolic disease in the host. Inflammation caused by infection leads to increased production of reactive oxygen species (ROS) and subsequent oxidative DNA damage. These in turn cause further inflammation and exacerbation of DNA damage, and pose a risk for cancer development. *Helicobacter pylori*-mediated inflammation has been implicated in gastric cancer in many previously established studies, and *Fusobacterium nucleatum* presence has been observed with greater intensity in colorectal cancer patients. Despite ambiguity in the exact mechanism, infection-mediated inflammation may have a link to cancer development through an accumulation of potentially mutagenic DNA damage in surrounding cells. The multiple DNA repair pathways such as base excision, nucleotide excision, and mismatch repair that are employed by cells are vital in the abatement of accumulated mutations that can lead to carcinogenesis. For this reason, understanding the role of DNA repair as an important cellular mechanism in combatting the development of cancer will be essential to characterizing the effect of infection on DNA repair proteins and to identifying early cancer biomarkers that may be targeted for cancer therapies and treatments.

**Keywords:** bacterial infection, commensal bacteria, DNA damage, inflammation and cancer, *Fusobacterium nucleatum*, *Helicobacter pylori*

## INTRODUCTION

The significance of cancer as a disease that affects a large percentage of the world population is undeniable. It is one of the leading causes of death worldwide and according to World Health Organization, it causes 8–9 million deaths/year. In the United States alone, it is projected that 39.6% of the population will have some type of cancer during their life. Consequently, there are enormous expenditures in the field of cancer care. The expenditures for cancer care in the United States were nearly \$125 billion in 2010 and could reach \$156 billion in 2020; as mentioned in the website of National Cancer Institute. Infection and infection-associated inflammation is the major threat of cancer. Chronic inflammation from infection causes abnormal immune response, obesity, DNA damage and cancer. The best example is the inflammatory bowel disease where Ulcerative colitis and Crohn's disease develop colon cancer.

In 1863, Rudolf Virchow was the first scientist to link inflammation with cancer. He mentioned that the origin of cancer was at sites of chronic inflammation and a group of irritants with tissue injury causes cell proliferation (Balkwill and Mantovani, 2001). How does inflammation

initiate malignancies? The possible answer to this question is the infection associated with chronic inflammation. Approximately 20% of cancer worldwide is caused by infection (Kuper et al., 2000; Parkin, 2006). The last 20 years of research shows that microbial infection is associated with cancer and can induce cancer progression. According to the estimate of the International Agency for Research on Cancer (IARC) approximately 18% of cancer are associated to infectious diseases that are caused by bacteria, viruses, and parasites. The well-known examples are human papilloma viruses (HPV; causing anogenital cancers), *Helicobacter pylori* (gastric cancers), hepatitis B and C viruses (hepatic cancers), and *Fusobacterium* (Colon cancer). The detailed list of microbes that have been researched in relation to their effect on certain DNA repair proteins are added in the **Table 1**. The gaps in knowledge in this field are reflected in the table, where there are many pathogen-associated cancers that have not been studied in relation to DNA repair proteins.

As cited in the press release of the Nobel Assembly (Marshall and Warren, 1984, 2005): “Many diseases in humans such as Crohn’s disease, ulcerative colitis, rheumatoid arthritis, and atherosclerosis are due to chronic inflammation. The discovery that one of the most common diseases of mankind, peptic ulcer disease, has a microbial cause has stimulated the search for microbes as possible causes of other chronic inflammatory conditions. Even though no definite answers are at hand, recent data clearly suggest that a dysfunction in the recognition of microbial products by the human immune system can result in disease development. The discovery of *Helicobacter pylori* has led to an increased understanding of the connection between chronic infection, inflammation, and cancer.”

Chronic infection generates a milieu of inflammatory cytokines that leads to inflammatory microenvironment, a critical modulator of carcinogenesis. The persistent infection and chronic inflammation changes somatic cells by the influence of associated microbes and epigenetic factors (Fernandes et al., 2015). Hanahan et al. showed that genome instability and inflammation are the emerging hallmarks associated with cancer (Hanahan and Weinberg, 2011). **Figure 1** shows the responsive elements that can trigger carcinogenesis. Bacterial infections increase cancer risk through either an extrinsic pathway, linked to induction of chronic inflammatory diseases that can increase cancer risk, or an intrinsic pathway, which is the accrual of genetic mutations that cause inflammation and transformation (Mantovani et al., 2008). Chronic inflammation has been associated with multiple types of cancer to the extent that inflammation period has been linked to increased risk of carcinogenesis (Shacter and Weitzman, 2002). Chronic inflammation is able to adjust the tumor microenvironment with cells such as tumor associated macrophages and various inflammatory agents such as chemokines, to regulate both tumor growth and angiogenesis (Coussens and Werb, 2002). Inflammation is also able to induce growth factors that serve several roles in carcinogenesis and tumorigenesis (Hanahan and Weinberg, 2011). The intrinsic pathway of genome alterations caused by infection is often linked to inflammation-mediated reactive oxygen species (ROS) production, which can increase

the rate of genetic mutations that can accumulate to cause cancer (Hanahan and Weinberg, 2011).

Bacterial infection causes inflammatory response and the ROS generated by bacterial infection often results in genomic instability (Chumduri et al., 2016). This implicates a bacterial infection in compromising or at least impeding some of the several cellular mechanisms for maintaining genetic integrity and repairing mutations. As genomic instability is an underlying factor in almost all cancer cells, the link between infection and cancer development and progression is a significant one. This review will focus on the mechanism of inflammation and ROS production post-infection, and then elaborate on the genomic instability induced by infection/inflammation by discussing the effect on various DNA repair pathways. As a major focus, we will bring the link of *Helicobacter pylori* with gastric cancer and microbial infection associated colorectal cancer under the scope of DNA damage repair following infection.

## MICROBIAL INFECTION-MEDIATED INFLAMMATION LINKED TO CANCER

Innate and adaptive immune responses are important to protect self from pathogenic microbial attack. Understanding of the infection process is important as bacterial and viral infection induces the inflammation that increases cancer risk (de Martel and Franceschi, 2009). The innate and adaptive immune responses will be discussed in the next section.

### Innate Immune Response

The immune response following recognition and invasion of microbes such as bacteria and viruses split into the innate and adaptive responses. Recognition and the initial precautionary actions are taken by the pattern recognition receptors (PRRs) of innate immune system (Pasare and Medzhitov, 2004; Akira et al., 2006). PRRs are found on the surfaces of epithelial cells and several immune cells to recognize structurally conserved pathogen-associated microbial products (PAMPs). For example, Toll-like receptor 4 (TLR4) is a PRR that binds lipopolysaccharide (LPS) on the outer membrane of gram-negative bacteria; TLR2 binds bacterial lipoproteins and lipoteichoic acids. The cytosolic sensor; NOD-like receptors (NLRs) detect intracellular pathogens. The innate immune response is the immediate mechanism by which the host attempts to clear an infection. Most of the PRRs lead to activation of MYD88-dependent pathways that eventually lead to NF- $\kappa$ B activation. This leads to further production of inflammatory cytokines and chemokines as well as antimicrobial peptides. Like TLRs, NLRs lead to an inflammatory response from both this signaling cascade and through the activation of caspases that act on cytokines that mediate the rest of the inflammatory response (Zarembek and Godowski, 2002; Barton and Medzhitov, 2003; Basset et al., 2003; Tsung et al., 2007; Church et al., 2008; Martinon et al., 2009; West et al., 2011b). The PRRs mediated inflammatory response is initiated by various inflammatory cytokines and chemokines, which draw macrophages and mast cells that



**TABLE 1** | A compilation of bacteria, virus, and parasite-associated cancers with some of the available information on their link to BER, NER, and MMR protein expression or mutations.

Bacteria-associated cancers	Gastric cancer	<i>Helicobacter pylori</i>		
		BER	NER	MMR
		APE1 (Ding et al., 2004)		MSH2 and MLH1 (Kim et al., 2002)
	Colon cancer	<i>Enteropathogenic Escherichia coli</i>		
		BER	NER	MMR
				MSH2 and MLH1 (Maddocks et al., 2013)
		<i>Streptococcus bovis</i> (Ellmerich et al., 2000)		
		BER	NER	MMR
		<i>Fusobacterium nucleatum</i> (Repass et al., 2016)		
		BER	NER	MMR
	Lung cancer	<i>Mycobacterium tuberculosis</i> (Wu et al., 2011)		
		BER	NER	MMR
		<i>Chlamydia pneumoniae</i> (Koyi et al., 2001)		
		BER	NER	MMR
	Bladder cancer	<i>Schistosoma hematobium</i> (Jemal et al., 2011)		
		BER	NER	MMR
		8-oxo-dG (Salim et al., 2008)		
		APE1 (Salim et al., 2008)		
Virus-associated cancers	Ovarian cancer	<i>Salmonella enterica</i> serovar Typhi (Dutta et al., 2000)		
		<i>Chlamydia trachomatis</i> (Xie et al., 2017)		
		BER	NER	MMR
		<i>Mycoplasma genitalium</i>		
	Cervical cancer	Human Papilloma Virus (HPV) (Bosch et al., 1995)		
		BER	NER	MMR
		XRCC1 (Bajpai et al., 2016)	ERCC2 (Bajpai et al., 2016)	hMLH1 (Ciavattini et al., 2005)
	Head and neck (Oropharyngeal) Squamous cell cancer	APE-1 (Shekari et al., 2008)	ERCC4 (Bajpai et al., 2016)	hMSH2 (Ciavattini et al., 2005)
				ERCC5 (Joo et al., 2016)
		hMLH3 (Ye et al., 2014)		
		Human Papilloma Virus (HPV) (Hajek et al., 2017)		
		BER	NER	MMR
		XRCC1 (Nickson et al., 2017)	ATR (Maddocks et al., 2013)	
		DNA polymerase $\beta$ (Nickson et al., 2017)	ERCC1 (Langer, 2012)	

(Continued)

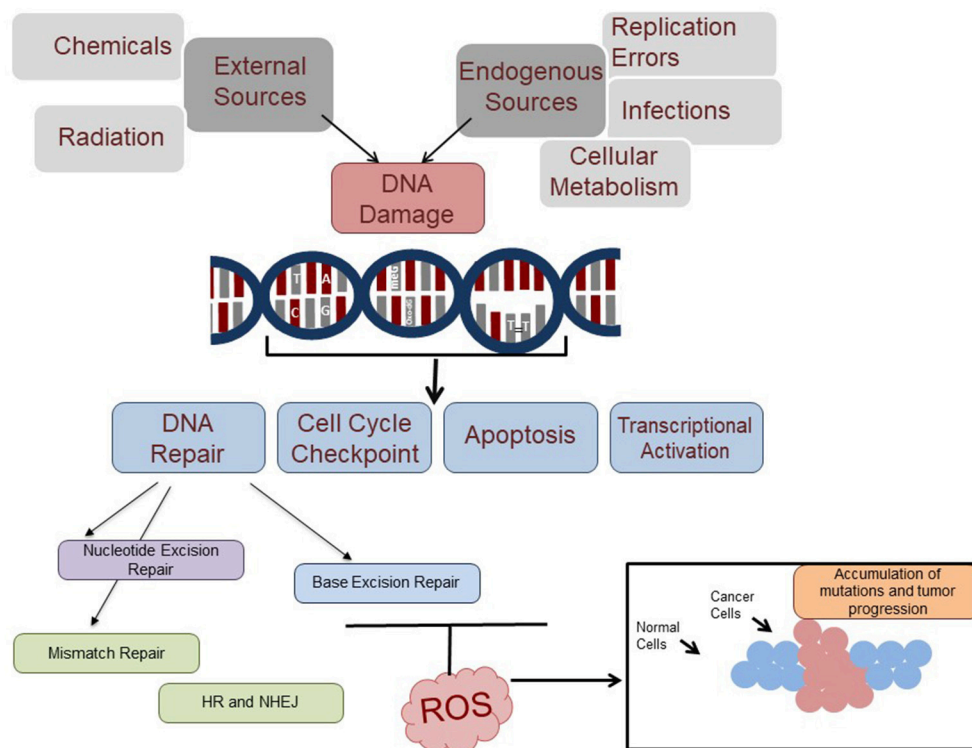
TABLE 1 | Continued

Hepatocellular Carcinoma	PNKP (Nickson et al., 2017)		
	PARP-1 (Nickson et al., 2017)		
	Hepatitis B Virus (HBV) (El-Serag, 2012; Lu et al., 2017)		
	BER	NER	MMR
	XRCC1 (Almeida Pereira Leite et al., 2013; Bose et al., 2013)	TFIIH (Arbuthnot and Kew, 2001; Qadri et al., 2011)	
	OGG1 (Bose et al., 2013)	TFIIB (Arbuthnot and Kew, 2001)	
	TDG (van de Klundert et al., 2012)	XPB (Jia et al., 1999; Arbuthnot and Kew, 2001)	
		XPB (Jia et al., 1999; Arbuthnot and Kew, 2001)	
		XAP-1 (DDB1) (Sohn et al., 2000)	
	Hepatitis C Virus (HCV) (El-Serag, 2012; Lu et al., 2017)		
	BER	NER	MMR
	Nei1 (Higgs et al., 2014)	XPB (Gulnaz et al., 2013)	hMSH2 (Helal et al., 2010)
	XRCC1 (Gulnaz et al., 2013)		hMLH1 (Helal et al., 2010)
	XRCC3 (Gulnaz et al., 2013)		
Nasopharyngeal Cancer	Epstein-Barr Virus (EBV) (Huang et al., 2017)		
	BER	NER	MMR
		XPA (Fu et al., 2015)	
T Cell leukemia	Human T Lymphotropic Virus Type I (HTLV-1) (Liao, 2006)		
	BER	NER	MMR
	XRCC5 (Ng et al., 2001)	ERCC5 (Ng et al., 2001)	
Kaposi sarcoma	Kaposi sarcoma-associated Herpes Virus (KSHV or HHV8) (Liao, 2006)		
	BER	NER	MMR
	APE1 (Zhong et al., 2017)		
	UNG2 (Verma et al., 2006)		
Parasite-associated cancer	Urinary bladder cancer	<i>Schistosoma haematobium</i> (Kawanishi et al., 2016)	
		BER	MMR
	Cholangio carcinoma	<i>Opisthorchis Viverrini</i> (Kawanishi et al., 2016)	
		BER	MMR
			hMSH2 (Liengswangwong et al., 2006)
			hMLH1 (Liengswangwong et al., 2006)

There are many gaps in knowledge in this field, particularly in the field of bacteria-associated cancers and regulation of various DNA repair pathway proteins.

release inflammatory mediators to recruit neutrophils and plasma proteins. Neutrophils phagocytose the pathogen and surrounding debris. Phagolysosomes form through the fusion of phagosomes with granules consisting of enzymes and ROS

that can kill the phagocytosed pathogen. The neutrophils can release these toxic granules and cause collateral damage to the surrounding tissue (Segal, 2005; Medzhitov, 2008; Serhan et al., 2008).



**FIGURE 1 |** DNA damage response and the potential role of ROS in inhibition of DNA repair. DNA damage is induced by external (various environmental pollutants, chemicals, and radiation), and internal resources (infection, cellular metabolism and replication errors). These DNA damages affect the cell cycle check point, apoptosis, transcriptional activation and cancer. Part of the DNA damages (DNA base adducts, mismatch bases, damaged bases, and double strand breaks) are repaired by the nucleotide excision repair, mismatch repair, base excision repair, and homologous recombination (HR) and non-homologous end joining (NHEJ) pathways, respectively. Many of the DNA damages are ROS-induced and recognized by the BER pathway which excises and repairs the lesions. However, ROS may potentially inhibit repair through down-regulation of certain initial proteins in the BER pathway, which can cause a buildup of carcinogenic mutations and ultimately lead to tumor progression.

The increase in the pro-inflammatory cytokines will clear the bacteria and the release of pro-inflammatory agents will be halted (**Figure 2**), when the tissue repair phase will be started. The tissue damage and debris resulting from the neutrophil activity may prevent impedance of inflammatory response and lead to continued chronic inflammation (Nathan, 2002). Various non-degradable components of the eliminated pathogen also contribute to a lasting inflammatory response even after the threat of the invading pathogen has been erased (Medzhitov, 2008).

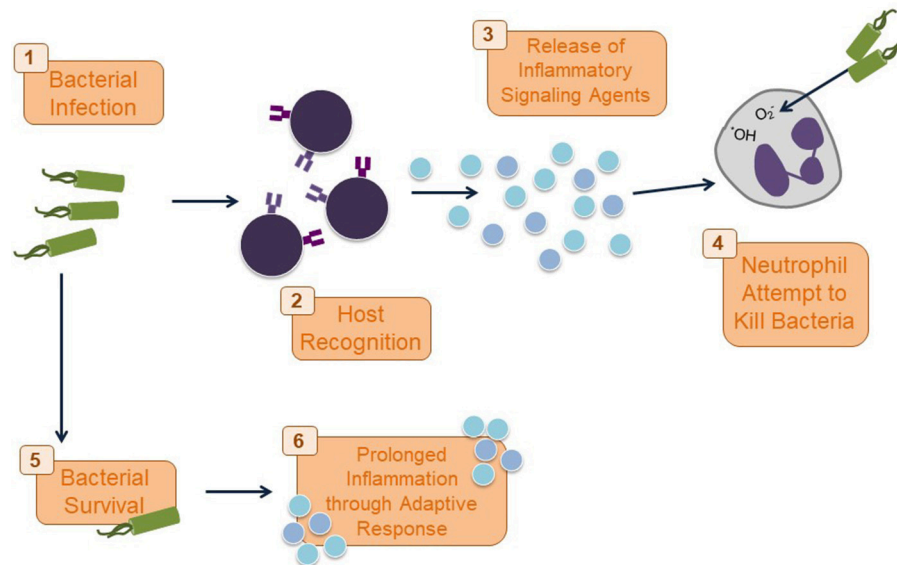
## Adaptive Immune Response

If the innate immune response is not adequate to kill the pathogen, the adaptive immune response is activated by continued inflammation at the site of infection (Zarembek and Godowski, 2002; Martinon et al., 2009). The microbes must then be cleared by specialized lymphocyte (B and T-cell)-mediated mechanisms (Medzhitov, 2008). For this reason, the adaptive immune response requires time to be initiated. Prolonged infection can lead to continuous tissue damage and a chronic inflammatory response that results in various diseases. For instance, in the intestinal tract, inflammatory

bowel diseases such as Crohn's disease and ulcerative colitis may manifest due to conditions of chronic inflammation that result from excessive immune activation exacerbating the initial inflammatory response to a foreign or commensal microbe (Cong et al., 2002; Macdonald and Monteleone, 2005). Specifically in Crohn's disease, the adaptive immune response plays a role through an extreme CD4T helper type I response to overexpression of various inflammatory cytokines (Macdonald and Monteleone, 2005).

## MECHANISMS THAT LINK INFLAMMATION TO CANCER PROGRESSION

The cytokines and chemokines from innate and adaptive immune cells direct the progression of the tumor microenvironment. Downstream effectors NF- $\kappa$ B, AP-1, STAT control the inflammatory milieu either by affecting tumor survival, growth, and tumor progression by signaling for molecules such as IL-6, IL-23 or by anti-tumor immunity (IFN- $\gamma$ , IL-12). The inflammatory cytokines have been reported to generate ROS and reactive nitrogen intermediates (RNI) using NADPH oxidase in phagocytic cells and epithelial cells (Yang et al., 2007). These



**FIGURE 2 |** Immune response to bacterial infection. Once a bacterial infection is recognized by PRRs on host cells as part of the innate immune response, inflammatory cytokines and chemokines are released to draw neutrophils to the site of infection. Neutrophils phagocytose and kill the bacteria with ROS. If this initial response does not kill the bacteria, the adaptive immune response kicks in and may result in a chronic inflammatory response.

ROS are the major source of damage to nucleic acid, proteins and lipids.

## Reactive Oxygen Species (ROS)

ROS refer to various, highly reactive and partially reduced metabolites of oxygen such as H<sub>2</sub>O<sub>2</sub> that are essential signaling molecules in the human immune system (Martindale and Holbrook, 2002; West et al., 2011b). ROS includes oxygen radicals (superoxide, hydroxyl, peroxy and alkoxy) and certain nonradicals that are either oxidizing agents and/or are easily converted into radicals, such as HOCl, ozone, peroxynitrite, singlet oxygen and hydrogen peroxide. ROS initiates DNA base oxidation which, if not repaired properly, may lead to induce a mutation.

Inflammatory cytokine signaling also depend on ROS. ROS is also crucial for inflammasome signaling and increased mitochondrial ROS activate the NLRP3 inflammasome (Zhou et al., 2011). Additionally, the post-translational modifications that ROS are associated with are often linked to modifications in protein cysteine residues that are generally associated with either Ca<sup>2+</sup> mediated signaling or tyrosine phosphorylation. This implicates ROS in cell motility, mitosis, differentiation, and immune response or regulation (van der Vliet, 2008). The immediate effect of excessive ROS presence in the host is a chronic inflammatory state that exacerbates both inflammation and, as a result, ROS production. Also, cancer cells utilize mROS to constitutively activate proliferation pathways to promote tumor growth (Cairns et al., 2011). Inability of cellular antioxidants to curtail ROS leads to oxidative stress on host cells that can lead to many adverse effects including induction of

DNA damage (Ernst, 1999). Therefore, following oxidative stress, the cell must survive by either adapting to the induced stress or by repairing the damage (Martindale and Holbrook, 2002). Inability to repair or adjust to the damage will lead to chronic conditions such as cancer, diabetes, and various neurological or cardiovascular diseases. The specific mechanisms of DNA damage repair that diminish the oxidative damage done by ROS will be discussed later in this review.

ROS in immune cells are important to kill extracellular pathogens by using the NADPH oxidase in phagosomes. Despite the various positive roles of ROS in immunity and other processes, they are associated with conditions such as diabetes, hypertension, cancer, and autoimmune diseases due to their ability to change and damage cellular proteins, lipids, and DNA (Zimmerman and Cerutti, 1984; Guzik et al., 2007). ROS is acquired through both exogenous and endogenous sources. Exogenous sources of ROS include carcinogen induced or generated ROS; for example, xenobiotics, chlorinated compounds, and radiation are associated to oxidative stress. The main endogenous source of ROS in the human body is through mitochondrial respiration or Ox-Phos (Oxidative phosphorylation) system. The electron transport chain uses mitochondrial oxidative phosphorylation complexes that lead to generation of ROS that can potentially harm cellular components such as proteins and nucleic acids through post-translational modifications and oxidation (Molina-Cruz et al., 2008). ROS produced by Ox-Phos pathway participate in immune signaling with TLR and cytosolic RIG-I like receptors (RLRs). West et al. reported that stimulation of cell-surface TLRs (TLR1, TLR2, and TLR4), but not



endosomal TLRs (TLR3, TLR7, TLR8, and TLR9), leads to an increase in mROS (Mitochondrial Reactive Oxygen Species) production through TRAF6 and ECSIT signaling (West et al., 2011a). Mitochondrial ROS enhance RLR signaling in autophagy-dependent Atg5-depleted cells that indicate the importance of autophagy in innate antiviral defense (Tal et al., 2009).

The most significant damages caused to DNA due to high concentrations of ROS include double and single strand breaks, oxidized DNA bases, and aberrant DNA cross-linking. Each type of ROS plays a different role in inflicting damage. For example, hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) directly induces DNA damage. The versatility of different ROS is reflected in the wide array of DNA damage that they can cause. In general, the hydroxyl molecule ( $\text{OH}^-$ ) is the most damaging form of ROS, but other forms like the oxygen molecule ( $\text{O}_2$ ),  $\text{RO}_2$ , and RO are also capable of different types of damage. For instance, while  $\text{OH}^-$  can react with all of the bases and the deoxyribose backbone of DNA,  $\text{O}_2$  preferentially targets guanine residues (Wiseman and Halliwell, 1996; Valko et al., 2007; Imlay, 2008). In a wider scope, ROS are generally likely to react with and damage DNA through oxidation, methylation, depurination, and deamination. The lesion most often found due to oxidative DNA damage, 8-hydroxyguanine (8-OHdG) or the nucleoside 8-hydroxydeoxyguanosine, is generally considered as marker of oxidative damage incurred by a cell. For instance, many studies examined for increased production of ROS culminating in an increased level of oxidative DNA damage by measuring 8-hydroxyguanine (Dandona et al., 1996; Farinati et al., 1998). In another study, it was observed that P53 acts as a defense against ROS-mediated DNA oxidation in various experimental conditions by detecting the presence of 8-oxodeoxyguanosine in the DNA (Sablina et al., 2005). The 8-hydroxydeoxyguanosine and other oxidized DNA lesions (8-oxo-adenine, thymine glycol, 5-hydroxy-deoxycytidine) have also been observed in many mutated oncogenes and tumor suppressor genes and these lesions are able to induce further neoplastic mutations in the DNA. The presence of high levels of 8-oxoguanine lesions, along with many other oxidative lesions, was shown in the DNA of tumorous tissues from many patients with different types of cancers (Klaunig and Kamendulis, 2004). This indicates a serious implication of oxidative stress and oxidative DNA damage in carcinogenesis (Figure 1).

## INVOLVEMENT OF MICROBES IN ROS-LINKED CARCINOGENESIS

The gut microbiota supplement a significant portion of the human metabolism, and the composition and activity of the microbiota plays a large role in susceptibility to metabolic diseases such as hyperglycemia, hyperlipidemia, insulin resistance, and obesity (Vijay-Kumar et al., 2010; Spencer et al., 2011). Microbes protect themselves from the ROS generated by host using an enzyme called superoxide dismutase (Sod), which is abundant in cells throughout the body. This enzyme attaches (binds) to molecules of copper and zinc to break

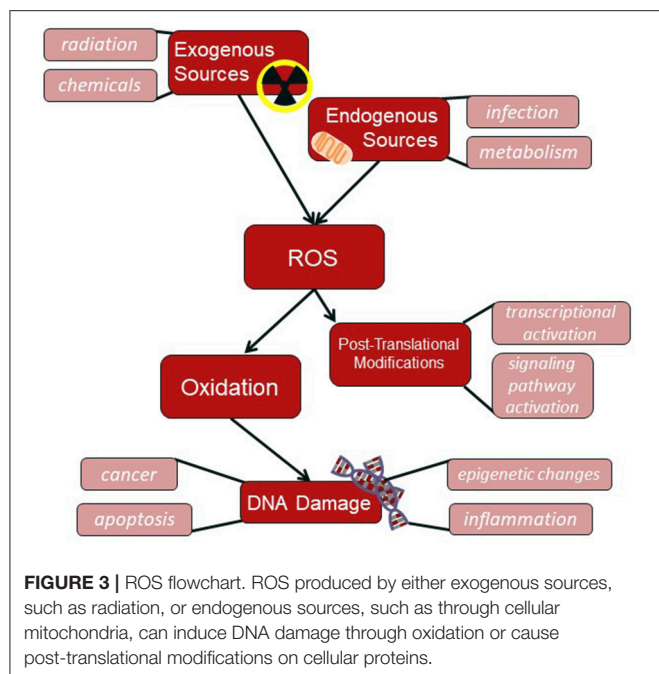
down toxic, charged oxygen molecules called superoxide radicals. Interestingly, bacteria has protective proteins such as SodA, SodB, SodC, AhpCF, KatG, KatE to detoxify ROS and proteins to counter damage (e.g., SoxRS, OxyRS, and SOS regulons) (Imlay, 2008). On the host side, Nox2, present in the NADPH complex is responsible for the generation of ROS (Lambeth, 2004). The generation of ROS is important for host defense as patients with chronic granulomatous disease (CGD) with deficiencies in NOX2 components are susceptible to infection (Cross et al., 2000).

Transcription factors vital to cellular processes such as inflammation, cell cycle regulation, motility, and growth like NF- $\kappa$ B, p53, HIF-1 $\alpha$ ,  $\beta$ -catenin/Wnt can be activated by the oxidative stress caused by an imbalance in the presence of oxidative ROS and the countering antioxidants. By mediating activity of genes and proteins related to oxidative stress, ROS are able to affect further cellular properties, such as cell growth, differentiation, and apoptosis, which can induce transformation (Reuter et al., 2010). For instance, ROS produced as a result of exposure of mouse mammary epithelial cells to MMP-3, a stromal enzyme that is linked to inducing epithelial-mesenchymal (EMT) transition and malignant transformation, caused activation of the transcription factor SNAIL, which induced oxidative DNA damage and EMT (Radisky et al., 2005). ROS-induced oxidative stress can also affect gene expression through direct DNA (de)methylation, which is an epigenetic method of silencing and activating certain genes by changing the physical accessibility of certain genes (Klaunig and Kamendulis, 2004). Chromosomal alterations induced by ROS can also lead to cellular damage. The genomic instability and transcriptional changes that accompany ROS and oxidative stress can therefore lead to carcinogenesis.

Post-infection inflammation-mediated mechanisms that assist in tumor formation and progression are an indirect method of bacterial infection leading to cancer. There are also several factors in a bacterial infection that can directly induce DNA damage or alter cell-signaling pathways that can lead to carcinogenesis. For instance, several bacteria such as *Escherichia coli* and *Shigella dysenteriae* are able to produce genotoxins (colibactin and shiga toxin, respectively) that inflict damage on the host DNA such as DNA strand breaks that may affect tumor suppressors or oncogenes (Gagnaire et al., 2017).

## DNA REPAIR MECHANISMS

Cellular DNA is altered either during replication or by external mutagens. Misincorporation of DNA bases can occur during replication; however, it is combatted through the proofreading activity of DNA polymerases. There may however be errors during DNA replication that are not recognized by the polymerases. There is also a great wealth of mutagens that can cause extensive changes in the sequence of human DNA. Of these, just oxidative DNA damage is estimated to arise about  $10^5$  times in 1 day due to ROS-induced damages (Lengauer et al., 1998). Accumulation of all these mutations would severely inhibit the ability of cells to survive and/or maintain proper cellular functions in almost all cases. Fortunately, the cell has several mechanisms that are specialized to recognize and repair



different types of DNA damage. Of these, this review will describe the mechanisms that address alterations in DNA sequence and their link to various cancers, including mismatch repair (MMR), nucleotide excision repair (NER), base excision repair (BER), and homologous recombination (HR) and non-homologous end joining (NHEJ). Most importantly, MMR, BER, HR, and NHEJ alterations have been linked to chronic inflammatory states (Figure 3). Therefore, the focus will be on genetic instabilities induced through deregulation of these repair pathways, in some cases due to inflammation, that lead to genetic instabilities that may contribute to cancer formation.

DNA repair pathways recognize and correct mismatches present in the DNA, abnormal bases, single-stranded and double-stranded DNA breaks (DSBs). The MMR, BER, and NER pathways respond to specific lesions in DNA residues. DSBs are particularly dangerous lesion and are repaired by two principal pathways: NHEJ pathway functions in all phases of the cell cycle, while the high-fidelity HR pathway requires a template for repair and utilizes available sister chromatids during the S and G2 phases of the cell cycle.

## Nucleotide Excision Repair

Nucleotide excision repair, or NER, is a versatile repair pathway that recognizes bulky adducts and general base lesions that cause a distortion of the double-helix structure of DNA. The major sources of these types of DNA damage are ultraviolet radiation and various types of genotoxic chemicals (Hoeijmakers, 2001). These lead to lesions such as pyrimidine dimers, cyclobutane pyrimidine dimers (CPD), and 6–4 photoproducts. NER-mediated pathway employs several different proteins to carry out a multi-step “cut-and-patch”-like pathway (Shuck et al., 2008). The defect in NER generates human genetic disorders and the bulky adducts targeted by NER mechanism can block

replication and/or transcription which can lead to apoptosis or necrosis (Sancar et al., 2004).

The NER mechanism is divided into two pathways that have different methods of lesion recognition. These are the global-genome (GG-NER) and transcription-coupled repair (TC-NER) pathways. In GG-NER, DNA damage is removed from the whole genome while TC-NER is primarily involved in repairing the damage on the coding strand of actively transcribed genes (Hoeijmakers, 2001; Jackson and Bartek, 2009). Both pathways differ in the initial recognition steps. In GG-NER, the major proteins involved in the recognition are the XPC/HR23B/CEN2 (XP complementation group C/Rad23 homolog B/Centrin-2) protein complex. TC-NER is important to protect the cells from UV-light-induced apoptosis. In TC-NER, the damage is recognized and RNA Pol II stalls, Cockayne syndrome protein CSB transiently interact with RNA Pol II and the other associated proteins can take care of the damage to repair. In patients with Cockayne syndrome have defective TC-NER (Sancar et al., 2004; Fousteri and Mullenders, 2008; Hanawalt and Spivak, 2008).

## Mismatch Repair

Mismatch repair, or MMR, fixes mismatched base pairs and insertion-deletion loops that are generally a product of incorrect genomic DNA replication. MMRs have the ability to inflict serious damage on the cell without killing it since they may go unnoticed and accumulate (Kunkel and Erie, 2005). MMR deficient cells can display a mutator phenotype, characterized by microsatellite instability and an elevated mutation frequency. The germline mutations in MMR genes can lead to a variety of cancers, including the non-polyposis colon cancer/ Lynch syndrome (Peltomäki, 2001). MMR involves three steps: a recognition step for identifying the mispaired bases, an excision step to remove the error-containing and the synthesis step, where the gap is filled-in by DNA polymerases. Therefore, the MMR pathway is very important to prevent cancer.

Several MMR proteins can be regulated upon chronic inflammation through the activation of HIF-1 $\alpha$  by inflammatory cytokines and ROS (Colotta et al., 2009). The inability of MMR to repair single base pair or small-scale mutations results in microsatellite instabilities such as poly CA repeats, present in various cancers (Lengauer et al., 1998).

## Base Excision Repair

Base-excision repair or BER recognizes a wide variety of damaged bases including those that underwent oxidation, alkylation, methylation, deamination, and hydroxylation (Hoeijmakers, 2001). It also repairs ROS induced strand breaks that consist of sugar fragments or 3'phosphate ends that are non-ligatable (Hegde et al., 2008). Due to its recognition and repair of an extensive range of ROS induced damages, BER is the major defense against accumulation of mutations caused by ROS. Additionally, the lesions targeted by BER are generally small-DNA base adducts. These types of lesions are more likely to be kept within the genome without cell apoptosis and can result in continued mutations in tumor suppressor and oncogenes that can lead to cancer (Hoeijmakers, 2001). Especially through this

framework, BER earns greater importance in preventing cellular transformation and cancer.

Single strand break repair pathway (SSBR) is now considered a specialized sub-pathway of BER. They share several common proteins including APE1, Pol $\beta$ , LIGIII $\alpha$ , along with the nick sensor poly (ADP-ribose) polymerase 1 (PARP1) and the scaffold protein X-ray cross-complementation group 1 (XRCC1) (Caldecott, 2008). ROS generate 8-oxoguanine (8-oxoG), ring-opened purines (formamidopyrimidines or Fapys), and other oxidized DNA base lesions that are repaired via the DNA BER pathway. BER begins with the recognition of an altered base by a DNA glycosylase. There are two classes of DNA glycosylases, the first group with the enzymes OGG1 and NTH1 utilize an internal Lys residue as the active site nucleophile. The second group comprising of NEIL1, NEIL2, and NEIL3 use N-terminal Pro or Val as the active site (Hazra et al., 2002, 2007; Sancar et al., 2004). The first and second group have distinct structural features and reaction mechanism but have overlapping substrate specificities. The NEIL proteins are able to preferentially target single stranded DNA and also lesions from a DNA bubble while NTH1 and OGG1 only excise lesions from double-stranded DNA, as they use the second strand as a template for repair (Hegde et al., 2008). Therefore, the NEIL proteins can be functional for replication/transcription errors (Banerjee et al., 2011).

These oxidized base-specific mammalian DNA glycosylases are bifunctional. Monofunctional DNA glycosylases excise the altered base in a way that leaves behind an AP site that needs to be processed by an AP endonuclease (3'-OH and 5'dRP are generated by APE1). On the other hand, the mammalian bifunctional DNA glycosylases have an associated AP lyase activity generating 3'dRP (OGG1, NTH1, and NEIL3) or 3'-P (NEIL1 and 2) and 5'-P. The DNA polymerase then fills in the gap using the template DNA strand and finally, DNA ligase seals the nick by completing the repair of the DNA duplex (Hitomi et al., 2007; Hegde et al., 2008; Maynard et al., 2009). The essential component of BER is the DNA glycosylase that recognizes and removes the oxidized base. Inhibition of the BER proteins may lead to an accumulation of oxidized DNA damage induced by ROS and the possible escalation of mutation rate upon an inflammatory response to bacterial infection that can contribute to carcinogenesis (Maynard et al., 2009).

The fact that defects in various components of NER and MMR have been shown to contribute to certain cancers indicates that the mutations that arise from inhibition of these repair mechanisms are certainly an important causative factor to cancer. For this reason, although BER has not been studied as thoroughly as NER and MMR in the context of cancer, it may have a significant role in the perpetuation of bacterial infection and inflammation mediated cancers such as gastric cancer and, to some degree, colorectal cancer (Wallace et al., 2012; Leguisamo et al., 2017). Various human DNA glycosylases have already been implicated in various types of cancer. NEIL2 was shown to protect against the oxidative damage that is induced by secondhand smoke in human lung cells, and lower levels of NEIL2 are associated with development of lung tumors (Sarker et al., 2014). It has also been shown that knock out of NEIL2

increased the accumulation of spontaneous mutations, and a variant of NEIL2 was observed in lung cancer samples (Dey et al., 2012). Furthermore, functional variants of NEIL2 have been linked to greater risk of squamous cell carcinoma in the oral cavity and oropharynx, making it a possible marker for risk to and progression of squamous cell carcinoma in the oral cavity and oropharynx (Zhai et al., 2008). In a study of human non-small cell lung cancer, it was found that carriers who were positive for silencing of the DNA glycosylase OGG1 via methylation had a 2.25-fold higher risk of developing non-small cell lung cancer than carriers who did not exhibit OGG1 methylation (Qin et al., 2017). In another study, an OGG1 variant (Ser326Cys polymorphism) was found to increase lung cancer risk by 24% in an analysis of seven studies totaling over 3,000 cases and controls. Both the mRNA and protein expression of MUTYH, a human DNA glycosylase that repairs the foremost oxidative DNA damage in prostate cancer (8-hydroxyguanine), was down-regulated in about two thirds of prostate cancers compared to the non-cancerous prostate tissue data presented in two separate publicly available databases (Shinmura et al., 2017). The apparent links between DNA glycosylase inhibition or silencing and various cancers indicates that: (1) Oxidative damage is a major contributor to the accumulation of genetic mutations that can lead to carcinogenesis, and (2) BER plays a significant role in repressing the accumulation of ROS-induced mutations and therefore inhibition of certain BER proteins may contribute to carcinogenesis. The roles of various bacterial infections in the development of gastric and colon cancer will be discussed later to explain the relationship of bacterial infection to DNA damage and repair inhibition.

## Homologous Recombination and Non-homologous End Joining

HR and NHEJ are mechanisms to repair double strand breaks in DNA (Jackson and Bartek, 2009). These repair processes are important since double strand breaks can be extremely harmful to the genome and are largely considered the most lethal type of DNA lesions since both strands of the DNA are affected (Helleday et al., 2008). Double strand breaks can be induced by X-rays, genotoxic chemicals, during replication of single strand breaks, or by ROS (Hoeijmakers, 2001). Due to the possibility of ROS inducing double strand breaks in the DNA, bacterial infection and the resulting inflammation are implicated in this type of DNA damage. HR is present during DNA replication in the S and G2 phase of the DNA cycle while NHEJ combats direct double strand breaks that can be induced by the other factors listed above, such as X-ray or ROS exposure, and is predominant in the G1 phase of the cell cycle (Hoeijmakers, 2001; Jackson and Bartek, 2009).

In HR, the MRN sensor complex; containing MRE11 (meiotic recombination 11), RAD50, and NBS1 recruits ATM (ataxia telangiectasia mutated), generates DNA breaks followed by phosphorylation of histone H2AX (generating  $\gamma$ H2AX) that amplifies the damage signal (Blackwood et al., 2013). Classical non-homologous end-joining (C-NHEJ) is the major pathway for DNA double strand break repair. Depletion of C-NHEJ factors



significantly abrogates double strand break repair in transcribed but not in non-transcribed genes (Chakraborty et al., 2016). In NHEJ, the Ku70-Ku80 initiates NHEJ, which are the sensor proteins that recruit DNA-PK (DNA-dependent protein kinase) and end-processing proteins, followed by ligation of the breaks by a complex consisting of DNA ligase IV/XRCC4; all the end-processing enzymes (Jackson and Bartek, 2009). Mutations in proteins that repair double strand breaks are linked to higher risk for different types of cancers, predominantly lymphomas (Hoeijmakers, 2001). Interestingly, HR occurs only in cycling cells while NHEJ occurs in all cells.

## NON-ROS LINKED INFLAMMATION-ASSOCIATED MECHANISMS FOR PROMOTING CARCINOGENESIS

Cancer related inflammation has been observed in most neoplastic tissues through observations of white blood cell and tumor-associated macrophage infiltration into the tumor microenvironment as well as pro-inflammatory cytokine and chemokine presence. These inflammatory factors have been linked to increased tissue remodeling and angiogenesis and are therefore stated to lead to cancer related inflammation (Balkwill and Mantovani, 2001; Colotta et al., 2009; Ostrand-Rosenberg and Sinha, 2009). Other than the intrinsic effect of cancer stimulating genomic instability induced by chronic inflammation, there are many components of an extrinsic inflammatory pathway that accelerate the further development of cancer. In this case, various pro-inflammatory cytokines, chemokines, and other inflammation-linked factors assist in tissue remodeling and angiogenesis.

For instance, inflammation activates NF- $\kappa$ B, which in turn activates various other inflammatory cytokines as well as angiogenic factors (Colotta et al., 2009). In the case of an acute inflammatory response of the innate immune system, the activation of NF- $\kappa$ B is not significant to the extent that it will affect angiogenesis and cancer progression. However, in the case of chronic inflammation, as is common in various bacterial infections such as *H. pylori* infection, the activation of these cytokines and factors will be consistent and significant enough to contribute to cancer development (Mantovani et al., 2008). For instance, the mediators that are downstream of NF- $\kappa$ B will help in neoplasia assisted by inflammation. Therefore, prolonged activation of NF- $\kappa$ B will promote tumor cell survival, initiation and progression of tumor tissue formation (Karin, 2006; Bollrath and Greten, 2009). In addition to NF- $\kappa$ B, various other inflammation-associated molecules such as IL-6 and TNF will aid in cancer progression through various mechanisms (Colotta et al., 2009). IL-6 specifically assists in tumor cell survival and growth. IL-6 is produced by myeloid-derived suppressor cells that, as suggested by the name, suppress T-cell activation. These cells are recruited to areas with chronic inflammation by pro-inflammatory mediators and are extremely influential in promoting cancer survival by allowing them to evade the immune system's attacks via T-cell activation.

TNF, on the other hand, mediates inflammation and can promote the growth of a tumor by assisting in angiogenesis, epithelial to mesenchymal transition, and other mechanisms. This pro-inflammatory cytokine is associated with tumor-associated macrophages, which are also found in areas of chronic inflammation. When tumor-associated macrophages secrete TNF, the activation of Wnt/ $\beta$ -catenin signaling pathway is promoted and this leads to greater tumor development (Colotta et al., 2009). The Wnt/ $\beta$ -catenin pathway is a significant signaling mechanism that controls transcription for proteins involved in cell proliferation and cell fate determination (MacDonald et al., 2009). Moreover, in both T cells and ECs, an upregulation of the Wnt/ $\beta$ -Catenin pathway occurs upon infection, a pathway usually associated with changes in the cellular turnover rate, tissue regeneration and cellular metabolism (Karin and Clevers, 2016). The activation of microbe-sensing pathways by ECs, associated with similar gene expression changes in both ECs and IELs in immune-response-related and metabolic pathways, pointed to ECs as potential primary microbe-responding cells that could prompt neighboring IELs. Lastly, NF- $\kappa$ B activation leads to the production of proangiogenic factors like vascular endothelial growth factor (VEGF), which allow for the tumor to grow and spread to distal sites to result in metastasis (Ellis and Hicklin, 2008). Of the countless links between inflammation and cancer progression, only a few have been discussed here. Ultimately, however, it is clear that cancer related inflammation is a significant factor in both the initiation of cancer as well as the consecutive growth and metastasis of a tumor.

## Bacterial Toxins Can Cause DNA Damage

Bacteria not only generates DNA damage but also interacts with the host DDR pathways so damage cannot be efficiently repaired. Some of the pathogenic strains of bacteria produce toxins such as Cytolethal distending toxin (cdt) and Colibactin. Cdt is present in *Campylobacter jejuni*, *Haemophilus ducreyi*, *Actinobacillus actinomycetemcomitans*, *Shigella dysenteriae*, *Helicobacter cinaedi*, *Helicobacter hepaticus*, *Salmonella* species. Cdt recruits the MRN complex (MRE11/Rad50/NBS1) and generate DSBs that ultimately progress to gastro-intestinal cancer (Taieb et al., 2016). Another toxin, Colibactin is a polyketide nonribosomal peptide produced by several species of *Enterobacteriaceae*; for example in some of the *E. coli* strain with pks, *Klebsiella pneumoniae* and *Enterobacter aerogenes*. Colibactin is responsible for alkylation and interstrand crosslinks of DNA followed by generation of DSBs (Nougayrède et al., 2006).

There are multiple mechanisms for bacterial infections to induce cancer, and we focused gastric cancer and colon cancer here. Further studies in this arena will give more in depth insight to the mechanisms of association between specific bacterial infections and cancers, but in the next section we will discuss the well-known and characterized bacterial-infection associated cancers especially through the lens of ROS induced DNA damage and alteration of DNA repair pathways.



## BACTERIAL INFECTION-ASSOCIATED WITH CANCERS

The cumulative effect of the various processes that take place after a bacterial infection, either through the direct action of the bacterium or indirectly through bacteria-induced pathways, is linked to carcinogenesis in various tissues. Many infectious agents have previously been linked to cancers, and are implicated in about 20% of human tumors (de Martel et al., 2012). For instance, respiratory tract and lung cancers have been linked to pulmonary infections caused by *Chlamydia pneumonia* and *Mycobacterium tuberculosis* (Chaturvedi et al., 2010). *Chlamydia trachomatis* and *Neisseria gonorrhoeae* infections have been associated with genitourinary cancers, and so on (Smith et al., 2004). Of the many bacterial species that have been implicated in cancer formation, *Helicobacter pylori* has been shown to play a significant role in contributing to the global gastric cancer burden and has therefore been studied more thoroughly than many other species linked to cancers. Despite a clear, demonstrated link between *H. pylori* infection and gastric cancer incidence, the exact mechanism of carcinogenesis has yet to be discovered and characterized in depth. *Fusobacterium nucleatum*, on the other hand, has been observed in large amounts in the intestinal tissues of colorectal cancer patients, but a link to cancer has not been concretely established as of yet (Gagnaire et al., 2017). Although there are some contributions available on the possible mechanisms with which it may be linked to colorectal cancer, they have not been properly defined either.

### Microbial Infection-Associated Mechanisms to Promote Cancer

In mammalian cells, all the above-mentioned repair mechanisms can repair damage using the DNA damage responses (DDRs) and failure in these bring DNA damage/mutation and genomic instability. Microbial infection is one of the major reasons of the failure of DDRs.

Microbes such as bacteria, virus and parasites are able to activate or alter various signaling pathways that may lead to either activation of oncogenes or down-regulation of tumor suppressor genes in a contribution to cancer progression (Francescone et al., 2014; Sheflin et al., 2014). Examples of these modifications and pathways are discussed in this section. They also modulate repair pathways as mentioned in the table that can generate mutations linked to cancer.

*Helicobacter pylori* infection leads to activation of PI3K-AKT pathway, which ultimately leads to degradation of tumor suppressor p53. It also contributes to cell transformation and growth by both preventing the degradation of and activating  $\beta$ -catenin through various bacterial effectors such as VacA (vacuolating cytotoxin A) (Tabassam et al., 2009). Similarly, a virulence factor of *Fusobacterium nucleatum* called *Fusobacterium* adhesin A (FadA) is able to bind E-cadherin to induce greater  $\beta$ -catenin release and ultimately activate WNT signaling, which is oncogenic (Rubinstein et al., 2013). *Helicobacter pylori* and *Salmonella enterica* serovar Typhimurium both activate MAPK and AKT signaling pathways

upon infection leading to altered mediation of cell growth, proliferation, migration, and other important processes relevant to cellular transformation (Sokolova et al., 2008; Gagnaire et al., 2017).

The bacterial toxins which cells are exposed when infected can alter the cell cycle and ultimately affect some of the processes, which are implicated in carcinogenesis: proliferation, apoptosis, and differentiation (Mager, 2006). Certain bacteria, which are classified as cyclomodulins, are able to change host cell cycle patterns with cell-cycle inhibitors, such as cytolethal distending toxins and cycle inhibiting factor, and cell cycle stimulators such as cytotoxic necrotizing factor (Nougayrède et al., 2005).

Bacteria of the microbiota are also implicated in cancer through their ability to construct biofilm. Biofilms form when bacteria aggregate and secrete a substance that allows them to stick to surfaces that generally have a mucosal lining (Johnson et al., 2016). In addition to being linked to inflammatory bowel conditions, bacterial biofilms have been seen on colorectal cancers, preferentially in proximal colon cancers, which have a higher mortality rate, than distal colon cancers (Dejea et al., 2014).

### *H. pylori* Associated Gastric Cancer

*Helicobacter pylori* is a gram-negative and spiral shaped bacterium. It has flagella that assist in movement and is able to survive at very low pH, making it a main colonizer of the stomach. Infection with *H. pylori* is associated with greater susceptibility to further infections, diarrhea, and chronic gastritis (Tomb et al., 1997; Crew and Neugut, 2006). While eradication of *H. pylori* infection is possible, there is high probability of relapse as well as antimicrobial resistance in many strains of the bacteria. The standard triple antibiotic treatment of *H. pylori* cures up to 70% of infected patients since there is a growing resistance to clarithromycin. Additionally, a very large portion of the world, about half of the total population, is exposed to or infected by *H. pylori* (Parsonnet et al., 1994; Malfertheiner et al., 2012). As *H. pylori* is the greatest risk factor for gastric cancer development, it is very important to study mechanisms of carcinogenesis upon infection in order to develop possible therapeutic and treatment options that address the specific pathways altered by *H. pylori*. Studies have shown that eradicating *H. pylori* infection decreases gastric cancer development in patients without premalignant tumors and prevents malignant transformation in patients with premalignant tumors (Wong et al., 2004; Malfertheiner et al., 2012). This further reinforces the link between infection and gastric cancer.

### Non-inflammatory Pathways for Cellular Transformation Upon *H. pylori* Infection

*H. pylori* has multiple bacterial effectors that are able to alter cellular signaling pathways in favor of carcinogenesis. For instance, vacuolating cytotoxin A (VacA) and outer inflammatory protein A (OipA) are involved in epidermal growth factor receptor activation which leads to PI3K-AKT signaling and ultimately activates  $\beta$ -catenin (Suzuki et al., 2009; Wroblewski et al., 2010). This signaling cascade leads

to transcriptional activation for cell growth. If the infection cannot be cleared, this becomes constitutively active and can then cause cellular transformation. This is just one example of how bacterial effectors can contribute to carcinogenesis in a non-inflammatory pathway. However, these effectors such as OipA are also able to induce proinflammatory cytokine expression along with other oncogenic proteins that have a significant effect on cellular transformation. Bacterial oncoproteins such as CagA are additional features of the bacteria that contribute to carcinogenesis (same reference as on line 573). The cag pathogenicity island, present in cag<sup>+</sup> strains of *H. pylori* has genes which encode for type IV bacterial secretion system, commonly known as T4SS, that is able to export bacterial proteins such as CagA upon bacterial attachment to host cells. Once it enters a host epithelial cell, this protein can be activated via phosphorylation to mitigate apoptosis and induce greater cell proliferation to contribute to carcinogenesis (Polk and Peek, 2010). Despite their active role in altering many cancer-associated cellular pathways, bacterial oncoproteins and effectors are only able to increase gastric cancer risk. The ultimate development of gastric cancer also relies on chronic inflammatory response to *H. pylori* infection, which is accompanied by a wide variety of consequences that contribute to cellular transformation (Lamb and Chen, 2013).

### Inflammation-Associated Pathways for Cellular Transformation Upon *H. pylori* Infection

The inflammatory response to *H. pylori* infection can significantly alter cellular signaling and activity to induce transformation through some of the many pathways already discussed in this review (Figueiredo et al., 2002). The focus here will be on DNA damage induced by *H. pylori* mediated inflammation as well as obstruction of DNA repair pathways by the bacterium.

*H. pylori* colonization is characterized by recurring infections even after assumed eradication of the bacteria as well as chronic gastritis, which signifies a continuous inflammatory response that is started as a host response to infection but is continued with the aid of the bacteria by effectors and oncoproteins that alter chemokine and cytokine release in the infected host cells (Miftahussurur et al., 2017). One such oncoprotein is Tipα, a membrane protein secreted by *H. pylori* that is associated with epithelial to mesenchymal transition by activating IL-6 cytokine-dependent STAT3 signaling and greatly impacts cancer cell invasiveness. Similar to IL-6, many other inflammatory cytokines play a role in extracellular matrix degradation to promote cell motility and angiogenesis (Chen et al., 2017).

An accumulation of DNA damage caused by chronic inflammation and resulting ROS has the potential to induce cellular transformation. In a normally functioning cell, the multiple DNA repair pathways are able to curb the accumulation of DNA mutations by addressing the damage as it occurs. In a cell infected by *H. pylori*, it is shown that mismatch repair (MMR), the major pathway that repairs small-scale mutations such as single base pair mismatches, is inhibited (Kim et al., 2002;

Santos et al., 2017). This potentially allows for seemingly minute mutations to accumulate in oncogenic or tumor suppressor genes and lead to cancer development. Two vital MMR proteins, MSH2 and MLH1, are directly affected by *H. pylori* infection. Additional MMR proteins MLH1, MSH3, MSH6, PMS1, and PMS2 have also been shown to be down-regulated upon *H. pylori* infection of gastric cell lines AGS and BG (Machado et al., 2013; Strickertsson et al., 2014; Santos et al., 2017). The various oxidative damages induced to DNA by *H. pylori* infection are supposed to be repaired by the BER pathway in a normally functioning cell. In a cell infected with *H. pylori*, there is decreased expression of vital BER proteins. APE-1, an AP endonuclease, and YB-1, an early-stage repair protein of BER, are down-regulated upon *H. pylori* infection (Machado et al., 2013). Down-regulation of the DNA glycosylase OGG1, which is very important for recognition and removal of abasic sites induced by bacterial and host ROS, has been observed in gastric epithelial cells upon infection as well. Abasic sites that are targeted by OGG1, such as 8oxodG lesions, are induced at a greater frequency upon *H. pylori* infection. Reduced expression of OGG1 allows for accumulation of abasic sites that would not be repaired by the other DNA repair pathways and lead to carcinogenic mutation-build up and cellular transformation (Kidane et al., 2014).

### Bacteria Associated Colon Cancer

The association between bacterial infection and colon cancer has not been elucidated to the extent that *H. pylori* induced gastric cancer has been. Certain pathogenic species such as enterotoxigenic *Bacteroides fragilis* and *Escherichia coli* strain NC101 have been linked to colitis-associated colon cancer (Wu et al., 2009; Arthur et al., 2012), but no bacterial species has been proven to be a major causative agent in colorectal carcinogenesis. Rather, a collection of gram-negative and anaerobic bacteria has been observed in colorectal tumor tissues and may serve as a marker of cancer. The colon is home to commensal bacterial species that play various supporting and valuable roles in processes such as metabolism in the host. These microbiota have also been implicated in carcinogenesis and tumor formation in the colon, potentially through bacterial dysbiosis in the gut (McCoy et al., 2013; Warren et al., 2013).

One area of interest in this field is the presence of *Fusobacterium nucleatum* in abundance in colorectal tumor tissues (Kostic et al., 2013). *F. nucleatum* is a gram-negative microbe more often associated with the oral cavity. However, the bacterial species have been identified in the early stages of cancer in colorectal adenomas as well as carcinoma samples (McCoy et al., 2013). Introduction of *F. nucleatum* to mice was shown to speed up colonic tumorigenesis and induce a pro-inflammatory state through NF-κB signaling (Kostic et al., 2013). *F. nucleatum* has also been positively correlated with mortality linked to colorectal cancer, meaning that greater abundance of the bacteria more likely resulted in mortality due to the cancer (Mima et al., 2016). There is still much to study concerning whether *F. nucleatum* is a causative agent in colorectal carcinogenesis, but the preliminary studies

have indicated the involvement of bacteria in the induction of cancer (Ray, 2011; Kostic et al., 2013; Gagnaire et al., 2017).

In the oral cavity, *F. nucleatum* is a very invasive bacterial species due to its ability to adhere well to mucous surfaces (McCoy et al., 2013). In the intestines, *F. nucleatum* can therefore act similar to *H. pylori* and adhere to host cell surfaces to alter cellular pathways with its bacterial proteins. Once *F. nucleatum* adheres to host cells, bacterial FadA adhesin binds E-cadherin to activate  $\beta$ -catenin signaling to increase cell growth and proliferation as well as to regulate inflammatory response of the cell (Rubinstein et al., 2013). *F. nucleatum* has also been found to activate TLR4 and ultimately lead to NF- $\kappa$ B activation (Yang et al., 2017). Many studies have linked *F. nucleatum* to a pro-inflammatory state (Kostic et al., 2013; Rubinstein et al., 2013). Inflammatory cytokine gene expression such as IL-10 and TNF- $\alpha$  has a positive association to abundance of *F. nucleatum* in the colon (Rubinstein et al., 2013). This also poses the possibility of *F. nucleatum* contributing to carcinogenesis and tumorigenesis by inducing genetic mutations as a result of prolonged inflammation and potentially the down-regulation of various DNA repair pathways, similar to what occurs in *H. pylori*-mediated inflammation leading to gastric cancer.

Other bacteria associated cancers: Other than *H. pylori* and *Fusobacterium*, *Salmonella typhi* infection has been associated with the development of gallbladder cancer (Mager, 2006; Di Domenico et al., 2017). All other bacteria associated with cancer is mentioned in the **Table 1**.

## BACTERIAL INFECTION IN CANCER PREVENTION AND THERAPY

While a number of bacterial infections have been shown to increase the potential for carcinogenesis, recent developments in the field have provided evidence for a positive role of certain bacteria and toxins in cancer prevention and therapy (Mager, 2006). For example in one case-control study, researchers found that *Helicobacter pylori* infection correlated to a lower risk for esophageal cancer development (de Martel et al., 2005). Alternatively, the introduction of bacteria or its toxins to treat cancer has become a method of interest for many types of cancer. Dr. William Coley started to treat end stage cancers with a vaccine made of killed *Streptococcus pyogenes* and *Serratia marcescens* in the late 1800's to induce an initial fever followed by treatment for many different types of cancers (de Martel et al., 2005).

More recently, researchers have shown that a vaccine with live attenuated *Salmonella enterica* serovar Typhi reduced tumor growth and enhanced survival in mice (Vendrell et al., 2011). The Bacillus Calmette-Guérin vaccine, which has a strain of *Mycobacterium bovis*, is used clinically for the treatment of high-risk urinary bladder cancer (Kucerova and Cervinkova, 2016). Finally, a number of bacteria species has been tried as anti-tumor agents in experimental models of cancer (Ryan et al., 2006).

## FUTURE RESEARCH DIRECTIONS

About 20% of the global cancer burden is linked to infectious agents including, but not limited to, *H. pylori*, *Hepatitis B* and *C* virus, and *Human papilloma virus* (Mantovani et al., 2008; Gagnaire et al., 2017). By studying how these infectious agents can lead to and exacerbate cancer states, we may be able to prevent certain cancers from forming or advancing and identify cancer markers or therapeutic targets in the treatment of cancer. It is already known that a chronic inflammatory state is able to induce DNA damage through oxidative stress induced by ROS production. As it is a hallmark of cancer, DNA damage must be repaired properly through the multiple machineries present within the cell (Colotta et al., 2009; Hanahan and Weinberg, 2011). However, along with inducing a chronic inflammatory state, bacterial infection may affect function and/or the level of DNA repair proteins leading to a buildup of genetic mutations in potential oncogenic and tumor suppressor genes that are major contributors to carcinogenesis. Studying the effect of infectious agents that are already linked to cancers, such as *F. nucleatum* and colorectal cancer, on the functionality of the various DNA repair pathways, can lead to novel identification of various markers for early cancer detection as well as more effective therapies and treatments that can combat the loss in DNA repair function in host cells.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

This work was supported, in whole or in part, by National Institute of Health Grants: DK107585 and DK099275 (to SD); R01 NS073976 (to TH).

## REFERENCES

- Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell* 124, 783–801. doi: 10.1016/j.cell.2006.02.015
- Almeida Pereira Leite, S. T., Marques-Guimarães, N., Silva-Oliveira, J. C., Dutra-Souto, F. J., Alves-dos-Santos, R., and Bassi-Branco, C. L. (2013). The X-ray repair cross complementing protein 1 (XRCC1) rs25487 polymorphism and susceptibility to cirrhosis in Brazilian patients with chronic viral hepatitis. *Ann. Hepatol.* 12, 733–739.
- Arbuthnot, P., and Kew, M. (2001). Hepatitis B virus and hepatocellular carcinoma. *Int. J. Exp. Pathol.* 82, 77–100. doi: 10.1111/j.1365-2613.2001.iep178.x
- Arthur, J. C., Perez-Chanona, E., Mühlbauer, M., Tomkovich, S., Uronis, J. M., Fan, T. J., et al. (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 338, 120–123. doi: 10.1126/science.1224820
- Bajpai, D., Banerjee, A., Pathak, S., Thakur, B., Jain, S. K., and Singh, N. (2016). Single nucleotide polymorphisms in the DNA repair genes



- in HPV-positive cervical cancer. *Eur. J. Cancer Prev.* 25, 224–231. doi: 10.1097/CEJ.0000000000000159
- Balkwill, F., and Mantovani, A. (2001). Inflammation and cancer: back to Virchow? *Lancet* 357, 539–545. doi: 10.1016/S0140-6736(00)04046-0
- Banerjee, D., Mandal, S. M., Das, A., Hegde, M. L., Das, S., Bhakat, K. K., et al. (2011). Preferential repair of oxidized base damage in the transcribed genes of mammalian cells. *J. Biol. Chem.* 286, 6006–6016. doi: 10.1074/jbc.M110.198796
- Barton, G. M., and Medzhitov, R. (2003). Toll-like receptor signaling pathways. *Science* 300, 1524–1525. doi: 10.1126/science.1085536
- Basset, C., Holton, J., O'Mahony, R., and Roitt, I. (2003). Innate immunity and pathogen-host interaction. *Vaccine* 21(Suppl. 2), S12–S23. doi: 10.1016/S0264-410X(03)00195-6
- Blackwood, J. K., Rzechorzek, N. J., Bray, S. M., Maman, J. D., Pellegrini, L., and Robinson, N. P. (2013). End-resection at DNA double-strand breaks in the three domains of life. *Biochem. Soc. Trans.* 41, 314–320. doi: 10.1042/BST20120307
- Bollrath, J., and Greten, F. R. (2009). IKK/NF-kappaB and STAT3 pathways: central signalling hubs in inflammation-mediated tumour promotion and metastasis. *EMBO Rep.* 10, 1314–1319. doi: 10.1038/embor.2009.243
- Bosch, F. X., Manos, M. M., Munoz, N., Sherman, M., Jansen, A. M., Peto, J., et al. (1995). Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J. Natl. Cancer Inst.* 87, 796–802. doi: 10.1097/00006254-199510000-00015
- Bose, S., Tripathi, D. M., Sukriti, Sakhuja, P., Kazim, S. N., and Sarin, S. K. (2013). Genetic polymorphisms of CYP2E1 and DNA repair genes HOGG1 and XRCC1: association with hepatitis B related advanced liver disease and cancer. *Gene* 519, 231–237. doi: 10.1016/j.gene.2013.02.025
- Cairns, R. A., Harris, I. S., and Mak, T. W. (2011). Regulation of cancer cell metabolism. *Nat. Rev. Cancer* 11, 85–95. doi: 10.1038/nrc2981
- Caldecott, K. W. (2008). Single-strand break repair and genetic disease. *Nat. Rev. Genet.* 9, 619–631. doi: 10.1038/nrg2380
- Chakraborty, A., Tapryal, N., Venkova, T., Horikoshi, N., Pandita, R. K., Sarker, A. H., et al. (2016). Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nat. Commun.* 7:13049. doi: 10.1038/ncomms13049
- Chaturvedi, A. K., Gaydos, C. A., Agreda, P., Holden, J. P., Chatterjee, N., Goedert, J. J., et al. (2010). *Chlamydia pneumoniae* infection and risk for lung cancer. *Cancer Epidemiol. Biomarkers Prev.* 19, 1498–1505. doi: 10.1158/1055-9965.EPI-09-1261
- Chen, G., Tang, N., Wang, C., Xiao, L., Yu, M., Zhao, L., et al. (2017). TNF-alpha-inducing protein of *Helicobacter pylori* induces epithelial-mesenchymal transition (EMT) in gastric cancer cells through activation of IL-6/STAT3 signaling pathway. *Biochem. Biophys. Res. Commun.* 484, 311–317. doi: 10.1016/j.bbrc.2017.01.110
- Chumduri, C., Gurumurthy, R. K., Zietlow, R., and Meyer, T. F. (2016). Subversion of host genome integrity by bacterial pathogens. *Nat. Rev. Mol. Cell Biol.* 17, 659–673. doi: 10.1038/nrm.2016.100
- Church, L. D., Cook, G. P., and McDermott, M. F. (2008). Primer: inflammasomes and interleukin 1beta in inflammatory disorders. *Nat. Clin. Pract. Rheumatol.* 4, 34–42. doi: 10.1038/nrcprheum0681
- Ciavattini, A., Piccioni, M., Tranquilli, A. L., Filosa, A., Pieramici, T., and Goteri, G. (2005). Immunohistochemical expression of DNA mismatch repair (MMR) system proteins (hMLH1, hMSH2) in cervical preinvasive and invasive lesions. *Pathol. Res. Pract.* 201, 21–25. doi: 10.1016/j.prp.2004.09.012
- Colotta, F., Allavena, P., Sica, A., Garlanda, C., and Mantovani, A. (2009). Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis* 30, 1073–1081. doi: 10.1093/carcin/bgp127
- Cong, Y., Weaver, C. T., Lazenby, A., and Elson, C. O. (2002). Bacterial-reactive T regulatory cells inhibit pathogenic immune responses to the enteric flora. *J. Immunol.* 169, 6112–6119. doi: 10.4049/jimmunol.169.11.6112
- Coussens, L. M., and Werb, Z. (2002). Inflammation and cancer. *Nature* 420, 860–867. doi: 10.1038/nature01322
- Crew, K. D., and Neugut, A. I. (2006). Epidemiology of gastric cancer. *World J. Gastroenterol.* 12, 354–362. doi: 10.3748/wjg.v12.i3.354
- Cross, A. R., Noack, D., Rae, J., Curmutte, J. T., and Heyworth, P. G. (2000). Hematologically important mutations: the autosomal recessive forms of chronic granulomatous disease (first update). *Blood Cells Mol. Dis.* 26, 561–565. doi: 10.1006/bcmd.2000.0333
- Dandona, P., Thustu, K., Cook, S., Snyder, B., Makowski, J., Armstrong, D., et al. (1996). Oxidative damage to DNA in diabetes mellitus. *Lancet* 347, 444–445. doi: 10.1016/S0140-6736(96)90013-6
- Dejea, C. M., Wick, E. C., Hechenbleikner, E. M., White, J. R., Mark Welch, J. L., Rossetti, B. J., et al. (2014). Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18321–18326. doi: 10.1073/pnas.1406199111
- de Martel, C., Ferlay, J., Franceschi, S., Vignat, J., Bray, F., Forman, D., et al. (2012). Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol.* 13, 607–615. doi: 10.1016/S1470-2045(12)70137-7
- de Martel, C., and Franceschi, S. (2009). Infections and cancer: established associations and new hypotheses. *Crit. Rev. Oncol. Hematol.* 70, 183–194. doi: 10.1016/j.critrevonc.2008.07.021
- de Martel, C., Llosa, A. E., Farr, S. M., Friedman, G. D., Vogelstein, J. H., Orentreich, N., et al. (2005). *Helicobacter pylori* infection and the risk of development of esophageal adenocarcinoma. *J. Infect. Dis.* 191, 761–767. doi: 10.1086/427659
- Dey, S., Maiti, A. K., Hegde, M. L., Hegde, P. M., Boldogh, I., Sarkar, P. S., et al. (2012). Increased risk of lung cancer associated with a functionally impaired polymorphic variant of the human DNA glycosylase NEIL2. *DNA Repair* 11, 570–578. doi: 10.1016/j.dnarep.2012.03.005
- Di Domenico, E. G., Cavallo, I., Pontone, M., Toma, L., and Ensoli, F. (2017). Biofilm producing *Salmonella typhi*: chronic colonization and development of gallbladder cancer. *Int. J. Mol. Sci.* 18:1887. doi: 10.3390/ijms18091887
- Ding, S. Z., O'Hara, A. M., Denning, T. L., Dirden-Kramer, B., Mifflin, R. C., Reyes, V. E., et al. (2004). *Helicobacter pylori* and H2O2 increase AP endonuclease-1/redox factor-1 expression in human gastric epithelial cells. *Gastroenterology* 127, 845–858. doi: 10.1053/j.gastro.2004.06.017
- Dutta, U., Garg, P. K., Kumar, R., and Tandon, R. K. (2000). Typhoid carriers among patients with gallstones are at increased risk for carcinoma of the gallbladder. *Am. J. Gastroenterol.* 95, 784–787. doi: 10.1111/j.1572-0241.2000.01860.x
- Ellis, L. M., and Hicklin, D. J. (2008). VEGF-targeted therapy: mechanisms of anti-tumour activity. *Nat. Rev. Cancer* 8, 579–591. doi: 10.1038/nrc2403
- Ellmerich, S., Schöller, M., Duranton, B., Gossé, F., Galluser, M., Klein, J. P., et al. (2000). Promotion of intestinal carcinogenesis by *Streptococcus bovis*. *Carcinogenesis* 21, 753–756. doi: 10.1093/carcin/21.4.753
- El-Serag, H. B. (2012). Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 142, 1264.e1–1273.e1. doi: 10.1053/j.gastro.2011.12.061
- Ernst, P. (1999). Review article: the role of inflammation in the pathogenesis of gastric cancer. *Aliment Pharmacol. Ther.* 13(Suppl. 1), 13–18. doi: 10.1046/j.1365-2036.1999.00003.x
- Farinati, F., Cardin, R., Degan, P., Rugge, M., Mario, F. D., Bonvicini, P., et al. (1998). Oxidative DNA damage accumulation in gastric carcinogenesis. *Gut* 42, 351–356. doi: 10.1136/gut.42.3.351
- Fernandes, J. V., DE Medeiros Fernandes, T. A., DE Azevedo, J. C., Cobucci, R. N., DE Carvalho, M. G., Andrade, V. S., et al. (2015). Link between chronic inflammation and human papillomavirus-induced carcinogenesis (Review). *Oncol. Lett.* 9, 1015–1026. doi: 10.3892/ol.2015.2884
- Figueiredo, C., Machado, J. C., Pharoah, P., Seruca, R., Sousa, S., Carvalho, R., et al. (2002). *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. *J. Natl. Cancer Inst.* 94, 1680–1687. doi: 10.1093/jnci/94.22.1680
- Fousteri, M., and Mullenders, L. H. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* 18, 73–84. doi: 10.1038/cr.2008.6
- Francescone, R., Hou, V., and Grivennikov, S. I. (2014). Microbiome, inflammation, and cancer. *Cancer J.* 20, 181–189. doi: 10.1097/PPO.0000000000000048
- Fu, X., Hu, J., Han, H. Y., Hua, Y. J., Zhou, L., Shuai, W. D., et al. (2015). High expression of XPA confers poor prognosis for nasopharyngeal carcinoma patients treated with platinum-based chemoradiotherapy. *Oncotarget* 6, 28478–28490. doi: 10.18632/oncotarget.4424



- Gagnaire, A., Nadel, B., Raoult, D., Neefjes, J., and Gorvel, J. P. (2017). Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nat. Rev. Microbiol.* 15, 109–128. doi: 10.1038/nrmicro.2016.171
- Gulnaz, A., Sayyed, A. H., Amin, F., Khan, A., Aslam, M. A., Shaikh, R. S., et al. (2013). Association of XRCC1, XRCC3, and XPD genetic polymorphism with an increased risk of hepatocellular carcinoma because of the hepatitis B and C virus. *Eur. J. Gastroenterol. Hepatol.* 25, 166–179. doi: 10.1097/MEG.0b013e328359a775
- Guzik, T. J., Hoch, N. E., Brown, K. A., McCann, L. A., Rahman, A., Dikalov, S., et al. (2007). Role of the T cell in the genesis of angiotensin II induced hypertension and vascular dysfunction. *J. Exp. Med.* 204, 2449–2460. doi: 10.1084/jem.20070657
- Hajek, M., Sewell, A., Kaech, S., Burtress, B., Yarbrough, W. G., and Issaeva, N. (2017). TRAF3/CYLD mutations identify a distinct subset of human papillomavirus-associated head and neck squamous cell carcinoma. *Cancer* 123, 1778–1790. doi: 10.1002/cncr.30570
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013
- Hanawalt, P. C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* 9, 958–970. doi: 10.1038/nrm2549
- Hazra, T. K., Das, A., Das, S., Choudhury, S., Kow, Y. W., and Roy, R. (2007). Oxidative DNA damage repair in mammalian cells: a new perspective. *DNA Repair* 6, 470–480. doi: 10.1016/j.dnarep.2006.10.011
- Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., et al. (2002). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3523–3528. doi: 10.1073/pnas.062053799
- Hegde, M. L., Hazra, T. K., and Mitra, S. (2008). Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.* 18, 27–47. doi: 10.1038/cr.2008.8
- Helal, T. E., Khamis, N. S., El-Sharkawy, T. M., Nada, O. H., and Radwan, N. A. (2010). Immunohistochemical expression of mismatch repair genes (hMSH2 and hMLH1) in hepatocellular carcinoma in Egypt. *APMIS* 118, 934–940. doi: 10.1111/j.1600-0463.2010.02658.x
- Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R. A. (2008). DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* 8, 193–204. doi: 10.1038/nrc2342
- Higgs, M. R., Chouteau, P., and Lerat, H. (2014). “Liver let die”: oxidative DNA damage and hepatotropic viruses. *J. Gen. Virol.* 95, 991–1004. doi: 10.1099/vir.0.059485-0
- Hitomi, K., Iwai, S., and Tainer, J. A. (2007). The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair* 6, 410–428. doi: 10.1016/j.dnarep.2006.10.004
- Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* 411, 366–374. doi: 10.1038/35077232
- Huang, J., Fogg, M., Wirth, L. J., Daley, H., Ritz, J., Posner, M. R., et al. (2017). Epstein-Barr virus-specific adoptive immunotherapy for recurrent, metastatic nasopharyngeal carcinoma. *Cancer* 123, 2642–2650. doi: 10.1002/cncr.30541
- Imlay, J. A. (2008). Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* 77, 755–776. doi: 10.1146/annurev.biochem.77.061606.161055
- Jackson, S. P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* 461, 1071–1078. doi: 10.1038/nature08467
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA Cancer J. Clin.* 61, 69–90. doi: 10.3322/caac.20107
- Jia, L., Wang, X. W., and Harris, C. C. (1999). Hepatitis B virus X protein inhibits nucleotide excision repair. *Int. J. Cancer* 80, 875–879. doi: 10.1002/(SICI)1097-0215(19990315)80:6<875::AID-IJC13>3.0.CO;2-Z
- Johnson, C. H., Spilker, M. E., Goetz, L., Peterson, S. N., and Siuzdak, G. (2016). Metabolite and microbiome interplay in cancer immunotherapy. *Cancer Res.* 76, 6146–6152. doi: 10.1158/0008-5472.CAN-16-0309
- Joo, J., Yoon, K. A., Hayashi, T., Kong, S. Y., Shin, H. J., Park, B., et al. (2016). Nucleotide excision repair gene ERCC2 and ERCC5 variants increase risk of uterine cervical cancer. *Cancer Res. Treat.* 48, 708–714. doi: 10.4143/crt.2015.098
- Karin, M. (2006). Nuclear factor-kappaB in cancer development and progression. *Nature* 441, 431–436. doi: 10.1038/nature04870
- Karin, M., and Clevers, H. (2016). Reparative inflammation takes charge of tissue regeneration. *Nature* 529, 307–315. doi: 10.1038/nature17039
- Kawanishi, S., Ohnishi, S., Ma, N., Hiraku, Y., Oikawa, S., and Murata, M. (2016). Nitrate and oxidative DNA damage in infection-related carcinogenesis in relation to cancer stem cells. *Genes Environ.* 38:26. doi: 10.1186/s41021-016-0055-7
- Kidane, D., Murphy, D. L., and Sweasy, J. B. (2014). Accumulation of abasic sites induces genomic instability in normal human gastric epithelial cells during *Helicobacter pylori* infection. *Oncogenesis* 3:e128. doi: 10.1038/oncsis.2014.42
- Kim, J. J., Tao, H., Carloni, E., Leung, W. K., Graham, D. Y., and Sepulveda, A. R. (2002). *Helicobacter pylori* impairs DNA mismatch repair in gastric epithelial cells. *Gastroenterology* 123, 542–553. doi: 10.1053/gast.2002.34751
- Klaunig, J. E., and Kamendulis, L. M. (2004). The role of oxidative stress in carcinogenesis. *Annu. Rev. Pharmacol. Toxicol.* 44, 239–267. doi: 10.1146/annurev.pharmtox.44.101802.121851
- Kostic, A. D., Chun, E., Robertson, L., Glickman, J. N., Gallini, C. A., Michaud, M., et al. (2013). *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 14, 207–215. doi: 10.1016/j.chom.2013.07.007
- Koyi, H., Brandén, E., Gnärpe, J., Gnärpe, H., and Steen, B. (2001). An association between chronic infection with *Chlamydia pneumoniae* and lung cancer. A prospective 2-year study. *APMIS* 109, 572–580. doi: 10.1034/j.1600-0463.2001.d01-177.x
- Kucerova, P., and Cervinkova, M. (2016). Spontaneous regression of tumour and the role of microbial infection—possibilities for cancer treatment. *Anticancer. Drugs* 27, 269–277. doi: 10.1097/CAD.0000000000000337
- Kunkel, T. A., and Erie, D. A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* 74, 681–710. doi: 10.1146/annurev.biochem.74.082803.133243
- Kuper, H., Adami, H. O., and Trichopoulos, D. (2000). Infections as a major preventable cause of human cancer. *J. Intern. Med.* 248, 171–183. doi: 10.1046/j.1365-2796.2000.00742.x
- Lamb, A., and Chen, L. F. (2013). Role of the *Helicobacter pylori*-induced inflammatory response in the development of gastric cancer. *J. Cell. Biochem.* 114, 491–497. doi: 10.1002/jcb.24389
- Lambeth, J. D. (2004). NOX enzymes and the biology of reactive oxygen. *Nat. Rev. Immunol.* 4, 181–189. doi: 10.1038/nri1312
- Langer, C. J. (2012). Exploring biomarkers in head and neck cancer. *Cancer* 118, 3882–3892. doi: 10.1002/cncr.26718
- Leguisamo, N. M., Gloria, H. C., Kalil, A. N., Martins, T. V., Azambuja, D. B., Meira, L. B., et al. (2017). Base excision repair imbalance in colorectal cancer has prognostic value and modulates response to chemotherapy. *Oncotarget* 8, 54199–54214. doi: 10.18632/oncotarget.14909
- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* 396, 643–649. doi: 10.1038/25292
- Liao, J. B. (2006). Viruses and human cancer. *Yale J. Biol. Med.* 79, 115–122.
- Liengswangwong, U., Karalak, A., Morishita, Y., Noguchi, M., Kuhaprema, T., Srivatanakul, P., et al. (2006). Immunohistochemical expression of mismatch repair genes: a screening tool for predicting mutator phenotype in liver fluke infection-associated intrahepatic cholangiocarcinoma. *World J. Gastroenterol.* 12, 3740–3745. doi: 10.3748/wjg.v12.i23.3740
- Lu, R., Zhang, Y. G., and Sun, J. (2017). STAT3 activation in infection and infection-associated cancer. *Mol. Cell. Endocrinol.* 451, 80–87. doi: 10.1016/j.mce.2017.02.023
- MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* 17, 9–26. doi: 10.1016/j.devcel.2009.06.016
- Macdonald, T. T., and Monteleone, G. (2005). Immunity, inflammation, and allergy in the gut. *Science* 307, 1920–1925. doi: 10.1126/science.1106442
- Machado, A. M., Desler, C., Bøggild, S., Strickertsson, J. A., Friis-Hansen, L., Figueiredo, C., et al. (2013). *Helicobacter pylori* infection affects mitochondrial function and DNA repair, thus, mediating genetic instability in gastric cells. *Mech. Ageing Dev.* 134, 460–466. doi: 10.1016/j.mad.2013.08.004
- Maddocks, O. D., Scanlon, K. M., and Sonnenberg, M. S. (2013). An *Escherichia coli* effector protein promotes host mutation via depletion of DNA mismatch repair proteins. *MBio* 4:e00152-13. doi: 10.1128/mBio.00152-13
- Mager, D. L. (2006). Bacteria and cancer: cause, coincidence or cure? A review. *J. Transl. Med.* 4:14. doi: 10.1186/1479-5876-4-14
- Malfertheiner, P., Megraud, F., O’Morain, C. A., Atherton, J., Axon, A. T., Bazzoli, F., et al. (2012). Management of *Helicobacter pylori*

- infection—the Maastricht IV/ Florence consensus report. *Gut* 61, 646–664. doi: 10.1136/gutjnl-2012-302084
- Mantovani, A., Allavena, P., Sica, A., and Balkwill, F. (2008). Cancer-related inflammation. *Nature* 454, 436–444. doi: 10.1038/nature07205
- Marshall, B. J., and Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1, 1311–1315. doi: 10.1016/S0140-6736(84)91816-6
- Marshall, B. J., and Warren, J. R. (2005). *The Bacterium Helicobacter pylori and Its Role in Gastritis and Peptic ulcer Disease*. The Nobel Prize in Physiology or Medicine, Press Release.
- Martindale, J. L., and Holbrook, N. J. (2002). Cellular response to oxidative stress: signaling for suicide and survival. *J. Cell. Physiol.* 192, 1–15. doi: 10.1002/jcp.10119
- Martinon, F., Mayor, A., and Tschopp, J. (2009). The inflammasomes: guardians of the body. *Annu. Rev. Immunol.* 27, 229–265. doi: 10.1146/annurev.immunol.021908.132715
- Maynard, S., Schurman, S. H., Harboe, C., de Souza-Pinto, N. C., and Bohr, V. A. (2009). Base excision repair of oxidative DNA damage and association with cancer and aging. *Carcinogenesis* 30, 2–10. doi: 10.1093/carcin/bgn250
- McCoy, A. N., Araujo-Perez, F., Azcarate-Peril, A., Yeh, J. J., Sandler, R. S., and Keku, T. O. (2013). *Fusobacterium* is associated with colorectal adenomas. *PLoS ONE* 8:e53653. doi: 10.1371/journal.pone.0053653
- Medzhitov, R. (2008). Origin and physiological roles of inflammation. *Nature* 454, 428–435. doi: 10.1038/nature07201
- Miftahussurur, M., Yamaoka, Y., and Graham, D. Y. (2017). *Helicobacter pylori* as an oncogenic pathogen, revisited. *Expert Rev. Mol. Med.* 19:e4. doi: 10.1017/erm.2017.4
- Mima, K., Nishihara, R., Qian, Z. R., Cao, Y., Sukawa, Y., Nowak, J. A., et al. (2016). *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut* 65, 1973–1980. doi: 10.1136/gutjnl-2015-310101
- Molina-Cruz, A., DeJong, R. J., Charles, B., Gupta, L., Kumar, S., Jaramillo-Gutierrez, G., et al. (2008). Reactive oxygen species modulate *Anopheles gambiae* immunity against bacteria and Plasmodium. *J. Biol. Chem.* 283, 3217–3223. doi: 10.1074/jbc.M705873200
- Nathan, C. (2002). Points of control in inflammation. *Nature* 420, 846–852. doi: 10.1038/nature01320
- Ng, P. W., Iha, H., Iwanaga, Y., Bittner, M., Chen, Y., Jiang, Y., et al. (2001). Genome-wide expression changes induced by HTLV-1 Tax: evidence for MLK-3 mixed lineage kinase involvement in tax-mediated NF-kappaB activation. *Oncogene* 20, 4484–4496. doi: 10.1038/sj.onc.1204513
- Nickson, C. M., Moori, P., Carter, R. J., Rubbi, C. P., and Parsons, J. L. (2017). Misregulation of DNA damage repair pathways in HPV-positive head and neck squamous cell carcinoma contributes to cellular radiosensitivity. *Oncotarget* 8, 29963–29975. doi: 10.18632/oncotarget.16265
- Nougayrède, J. P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., et al. (2006). *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313, 848–851. doi: 10.1126/science.1127059
- Nougayrède, J. P., Taieb, F., De Rycke, J., and Oswald, E. (2005). Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle. *Trends Microbiol.* 13, 103–110. doi: 10.1016/j.tim.2005.01.002
- Ostrand-Rosenberg, S., and Sinha, P. (2009). Myeloid-derived suppressor cells: linking inflammation and cancer. *J. Immunol.* 182, 4499–4506. doi: 10.4049/jimmunol.0802740
- Parkin, D. M. (2006). The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* 118, 3030–3044. doi: 10.1002/ijc.21731
- Parsonnet, J., Hansen, S., Rodriguez, L., Gelb, A. B., Warnke, R. A., Jellum, E., et al. (1994). *Helicobacter pylori* infection and gastric lymphoma. *N. Engl. J. Med.* 330, 1267–1271. doi: 10.1056/NEJM199405053301803
- Pasare, C., and Medzhitov, R. (2004). Toll-like receptors and acquired immunity. *Semin. Immunol.* 16, 23–26. doi: 10.1016/j.smim.2003.10.006
- Peltomäki, P. (2001). Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum. Mol. Genet.* 10, 735–740. doi: 10.1093/hmg/10.7.735
- Polk, D. B., and Peek, R. M. Jr. (2010). *Helicobacter pylori*: gastric cancer and beyond. *Nat. Rev. Cancer* 10, 403–414. doi: 10.1038/nrc2857
- Qadri, I., Fatima, K., and Abdel Hafiz, H. (2011). Hepatitis B virus X protein impedes the DNA repair via its association with transcription factor, TFIIF. *BMC Microbiol.* 11:48. doi: 10.1186/1471-2180-11-48
- Qin, H., Zhu, J., Zeng, Y., Du, W., Shen, D., Lei, Z., et al. (2017). Aberrant promoter methylation of hOGG1 may be associated with increased risk of non-small cell lung cancer. *Oncotarget* 8, 8330–8341. doi: 10.18632/oncotarget.14177
- Radisky, D. C., Levy, D. D., Littlepage, L. E., Liu, H., Nelson, C. M., Fata, J. E., et al. (2005). Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* 436, 123–127. doi: 10.1038/nature03688
- Ray, K. (2011). Colorectal cancer: *Fusobacterium nucleatum* found in colon cancer tissue—could an infection cause colorectal cancer? *Nat. Rev. Gastroenterol. Hepatol.* 8:662. doi: 10.1038/nrgastro.2011.208
- Repas, J., Maherali, N., Owen, K., and Reproducibility Project: Cancer Biology. (2016). Registered report: *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Elife* 5:e10012. doi: 10.7554/eLife.10012
- Reuter, S., Gupta, S. C., Chaturvedi, M. M., and Aggarwal, B. B. (2010). Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic. Biol. Med.* 49, 1603–1616. doi: 10.1016/j.freeradbiomed.2010.09.006
- Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y. W. (2013). *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe* 14, 195–206. doi: 10.1016/j.chom.2013.07.012
- Ryan, R. M., Green, J., and Lewis, C. E. (2006). Use of bacteria in anti-cancer therapies. *Bioessays* 28, 84–94. doi: 10.1002/bies.20336
- Sablina, A. A., Budanov, A. V., Ilyinskaya, G. V., Agapova, L. S., Kravchenko, J. E., and Chumakov, P. M. (2005). The antioxidant function of the p53 tumor suppressor. *Nat. Med.* 11, 1306–1313. doi: 10.1038/nm1320
- Salim, E. I., Morimura, K., Menesi, A., El-Lity, M., Fukushima, S., and Wanibuchi, H. (2008). Elevated oxidative stress and DNA damage and repair levels in urinary bladder carcinomas associated with schistosomiasis. *Int. J. Cancer* 123, 601–608. doi: 10.1002/ijc.23547
- Sancar, A., Lindsey-Boltz, L. A., Unsal-Kaçmaz, K., and Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 73, 39–85. doi: 10.1146/annurev.biochem.73.011303.073723
- Santos, J. C., Brianti, M. T., Almeida, V. R., Ortega, M. M., Fischer, W., Haas, R., et al. (2017). *Helicobacter pylori* infection modulates the expression of miRNAs associated with DNA mismatch repair pathway. *Mol. Carcinog.* 56, 1372–1379. doi: 10.1002/mc.22590
- Sarker, A. H., Chatterjee, A., Williams, M., Lin, S., Havel, C., Jacob, P. 3rd., et al. (2014). NEIL2 protects against oxidative DNA damage induced by sidestream smoke in human cells. *PLoS ONE* 9:e90261. doi: 10.1371/journal.pone.0090261
- Segal, A. W. (2005). How neutrophils kill microbes. *Annu. Rev. Immunol.* 23, 197–223. doi: 10.1146/annurev.immunol.23.021704.115653
- Serhan, C. N., Chiang, N., and van Dyke, T. E. (2008). Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nat. Rev. Immunol.* 8, 349–361. doi: 10.1038/nri2294
- Shacter, E., and Weitzman, S. A. (2002). Chronic inflammation and cancer. *Oncology* 16, 217–226, 229. discussion: 230–232.
- Shelton, A. M., Whitney, A. K., and Weir, T. L. (2014). Cancer-promoting effects of microbial dysbiosis. *Curr. Oncol. Rep.* 16:406. doi: 10.1007/s11912-014-0406-0
- Shekari, M., Sobti, R. C., Tamandani, D. M., Malekzadeh, K., Kaur, P., and Suri, V. (2008). Association of genetic polymorphism of the DNA base excision repair gene (APE-1 Asp148 Glu) and HPV type (16/18) with the risk of cervix cancer in north Indian population. *Cancer Biomark.* 4, 63–71. doi: 10.3233/CBM-2008-4202
- Shimura, K., Kato, H., Kawanishi, Y., Yoshimura, K., Igarashi, H., Goto, M., et al. (2017). Reduced expression of the DNA glycosylase gene MUTHYH is associated with an increased number of somatic mutations via a reduction in the DNA repair capacity in prostate adenocarcinoma. *Mol. Carcinog.* 56, 781–788. doi: 10.1002/mc.22509
- Shuck, S. C., Short, E. A., and Turchi, J. J. (2008). Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res.* 18, 64–72. doi: 10.1038/cr.2008.2
- Smith, J. S., Bosetti, C., Muñoz, N., Herrero, R., Bosch, F. X., Eluf-Neto, J., et al. (2004). *Chlamydia trachomatis* and invasive cervical cancer: a pooled analysis of the IARC multicentric case-control study. *Int. J. Cancer* 111, 431–439. doi: 10.1002/ijc.20257
- Sohn, S., Jaitovich-Groisman, I., Benlimame, N., Galipeau, J., Batist, G., and Alaoui-Jamali, M. A. (2000). Retroviral expression of the hepatitis B virus

- x gene promotes liver cell susceptibility to carcinogen-induced site specific mutagenesis. *Mutat. Res.* 460, 17–28. doi: 10.1016/S0921-8777(00)00010-0
- Sokolova, O., Bozko, P. M., and Naumann, M. (2008). *Helicobacter pylori* suppresses glycogen synthase kinase 3 $\beta$  to promote beta-catenin activity. *J. Biol. Chem.* 283, 29367–29374. doi: 10.1074/jbc.M801818200
- Spencer, M. D., Hamp, T. J., Reid, R. W., Fischer, L. M., Zeisel, S. H., and Fodor, A. A. (2011). Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* 140, 976–986. doi: 10.1053/j.gastro.2010.11.049
- Strickertsson, J. A., Desler, C., and Rasmussen, L. J. (2014). Impact of bacterial infections on aging and cancer: impairment of DNA repair and mitochondrial function of host cells. *Exp. Gerontol.* 56, 164–174. doi: 10.1016/j.exger.2014.03.024
- Suzuki, M., Mimuro, H., Kiga, K., Fukumatsu, M., Ishijima, N., Morikawa, H., et al. (2009). *Helicobacter pylori* CagA phosphorylation-independent function in epithelial proliferation and inflammation. *Cell Host Microbe* 5, 23–34. doi: 10.1016/j.chom.2008.11.010
- Tabassam, F. H., Graham, D. Y., and Yamaoka, Y. (2009). *Helicobacter pylori* activate epidermal growth factor receptor- and phosphatidylinositol 3-OH kinase-dependent Akt and glycogen synthase kinase 3 $\beta$  phosphorylation. *Cell. Microbiol.* 11, 70–82. doi: 10.1111/j.1462-5822.2008.01237.x
- Taieb, F., Petit, C., Nougayrède, J. P., and Oswald, E. (2016). The enterobacterial genotoxins: cytolethal distending toxin and colibactin. *EcoSal Plus* 7, 1–21. doi: 10.1128/ecosalplus.ESP-0008-2016
- Tal, M. C., Sasai, M., Lee, H. K., Yordy, B., Shadel, G. S., and Iwasaki, A. (2009). Absence of autophagy results in reactive oxygen species-dependent amplification of RLR signaling. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2770–2775. doi: 10.1073/pnas.0807694106
- Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., et al. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388, 539–547. doi: 10.1038/41483
- Tsung, A., Klune, J. R., Zhang, K., Jeyabalan, G., Cao, Z., Peng, X., et al. (2007). HMGB1 release induced by liver ischemia involves Toll-like receptor 4 dependent reactive oxygen species production and calcium-mediated signaling. *J. Exp. Med.* 204, 2913–2923. doi: 10.1084/jem.20070247
- Valko, M., Leibfritz, D., Moncol, J., Cronin, M. T., Mazur, M., and Telser, J. (2007). Free radicals and antioxidants in normal physiological functions and human disease. *Int. J. Biochem. Cell Biol.* 39, 44–84. doi: 10.1016/j.biocel.2006.07.001
- van de Klundert, M. A., van Hemert, F. J., Zaaier, H. L., and Kootstra, N. A. (2012). The hepatitis B virus x protein inhibits thymine DNA glycosylase initiated base excision repair. *PLoS ONE* 7:e48940. doi: 10.1371/journal.pone.0048940
- van der Vliet, A. (2008). NADPH oxidases in lung biology and pathology: host defense enzymes, and more. *Free Radic. Biol. Med.* 44, 938–955. doi: 10.1016/j.freeradbiomed.2007.11.016
- Vendrell, A., Gravisaco, M. J., Pasetti, M. F., Croci, M., Colombo, L., Rodriguez, C., et al. (2011). A novel *Salmonella typhi*-based immunotherapy promotes tumor killing via an antitumor Th1-type cellular immune response and neutrophil activation in a mouse model of breast cancer. *Vaccine* 29, 728–736. doi: 10.1016/j.vaccine.2010.11.017
- Verma, S. C., Bajaj, B. G., Cai, Q., Si, H., Seelhammer, T., and Robertson, E. S. (2006). Latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus recruits uracil DNA glycosylase 2 at the terminal repeats and is important for latent persistence of the virus. *J. Virol.* 80, 11178–11190. doi: 10.1128/JVI.01334-06
- Vijay-Kumar, M., Aitken, J. D., Carvalho, F. A., Cullender, T. C., Mwangi, S., Srinivasan, S., et al. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* 328, 228–231. doi: 10.1126/science.1179721
- Wallace, S. S., Murphy, D. L., and Sweasy, J. B. (2012). Base excision repair and cancer. *Cancer Lett.* 327, 73–89. doi: 10.1016/j.canlet.2011.12.038
- Warren, R. L., Freeman, D. J., Pleasance, S., Watson, P., Moore, R. A., Cochrane, K., et al. (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* 1:16. doi: 10.1186/2049-2618-1-16
- West, A. P., Brodsky, I. E., Rahner, C., Woo, D. K., Erdjument-Bromage, H., Tempst, P., et al. (2011a). TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* 472, 476–480. doi: 10.1038/nature09973
- West, A. P., Shadel, G. S., and Ghosh, S. (2011b). Mitochondria in innate immune responses. *Nat. Rev. Immunol.* 11, 389–402. doi: 10.1038/nri2975
- Wiseman, H., and Halliwell, B. (1996). Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. *Biochem. J.* 313 (Pt 1), 17–29. doi: 10.1042/bj3130017
- Wong, B. C., Lam, S. K., Wong, W. M., Chen, J. S., Zheng, T. T., Feng, R. E., et al. (2004). *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA* 291, 187–194. doi: 10.1001/jama.291.2.187
- Wroblewski, L. E., Peek, R. M. Jr., and Wilson, K. T. (2010). *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin. Microbiol. Rev.* 23, 713–739. doi: 10.1128/CMR.00011-10
- Wu, C. Y., Hu, H. Y., Pu, C. Y., Huang, N., Shen, H. C., Li, C. P., et al. (2011). Pulmonary tuberculosis increases the risk of lung cancer: a population-based cohort study. *Cancer* 117, 618–624. doi: 10.1002/cncr.25616
- Wu, S., Rhee, K. J., Albesiano, E., Rabizadeh, S., Wu, X., Yen, H. R., et al. (2009). A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* 15, 1016–1022. doi: 10.1038/nm.2015
- Xie, X., Yang, M., Ding, Y., and Chen, J. (2017). Microbial infection, inflammation and epithelial ovarian cancer. *Oncol. Lett.* 14, 1911–1919. doi: 10.3892/ol.2017.6388
- Yang, D., Elner, S. G., Bian, Z. M., Till, G. O., Petty, H. R., and Elner, V. M. (2007). Pro-inflammatory cytokines increase reactive oxygen species through mitochondria and NADPH oxidase in cultured RPE cells. *Exp. Eye Res.* 85, 462–472. doi: 10.1016/j.exer.2007.06.013
- Yang, Y., Weng, W., Peng, J., Hong, L., Yang, L., Toiyama, Y., et al. (2017). *Fusobacterium nucleatum* increases proliferation of colorectal cancer cells and tumor development in mice by activating toll-like receptor 4 signaling to nuclear factor-kappaB, and up-regulating expression of microRNA-21. *Gastroenterology* 152, 851.e24–866.e24. doi: 10.1053/j.gastro.2016.11.018
- Ye, F., Cheng, Q., Shen, J., Zhou, C., and Chen, H. (2014). Mismatch repair gene MLH3 Pro844Leu and Thr942Ile polymorphisms and the susceptibility to cervical carcinoma and HPV infection: a case-control study in a Chinese population. *PLoS ONE* 9:e96224. doi: 10.1371/journal.pone.0096224
- Zarembek, K. A., and Godowski, P. J. (2002). Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J. Immunol.* 168, 554–561. doi: 10.4049/jimmunol.168.2.554
- Zhai, X., Zhao, H., Liu, Z., Wang, L. E., El-Naggar, A. K., Sturgis, E. M., et al. (2008). Functional variants of the NEIL1 and NEIL2 genes and risk and progression of squamous cell carcinoma of the oral cavity and oropharynx. *Clin. Cancer Res.* 14, 4345–4352. doi: 10.1158/1078-0432.CCR-07-5282
- Zhong, C., Xu, M., Wang, Y., Xu, J., and Yuan, Y. (2017). An APE1 inhibitor reveals critical roles of the redox function of APE1 in KSHV replication and pathogenic phenotypes. *PLoS Pathog.* 13:e1006289. doi: 10.1371/journal.ppat.1006289
- Zhou, R., Yazdi, A. S., Menu, P., and Tschopp, J. (2011). A role for mitochondria in NLRP3 inflammasome activation. *Nature* 469, 221–225. doi: 10.1038/nature09663
- Zimmerman, R., and Cerutti, P. (1984). Active oxygen acts as a promoter of transformation in mouse embryo C3H/10T1/2/C18 fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.* 81, 2085–2087. doi: 10.1073/pnas.81.7.2085

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Sahan, Hazra and Das. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Targeting the Bacterial Orisome in the Search for New Antibiotics

Julia E. Grimwade\* and Alan C. Leonard

Department of Biological Sciences, Florida Institute of Technology, Melbourne, FL, United States

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
Fox Chase Cancer Center,  
United States

### Reviewed by:

Dhruba Chattoraj,  
National Institutes of Health (NIH),  
United States  
Anders Løbner-Olesen,  
University of Copenhagen, Denmark

### \*Correspondence:

Julia E. Grimwade  
grimwade@fit.edu

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 29 September 2017

**Accepted:** 15 November 2017

**Published:** 27 November 2017

### Citation:

Grimwade JE and Leonard AC  
(2017) Targeting the Bacterial  
Orisome in the Search for New  
Antibiotics. *Front. Microbiol.* 8:2352.  
doi: 10.3389/fmicb.2017.02352

There is an urgent need for new antibiotics to combat drug resistant bacteria. Existing antibiotics act on only a small number of proteins and pathways in bacterial cells, and it seems logical that expansion of the target set could lead to development of novel antimicrobial agents. One essential process, not yet exploited for antibiotic discovery, is the initiation stage of chromosome replication, mediated by the bacterial orisome. In all bacteria, orisomes assemble when the initiator protein, DnaA, as well as accessory proteins, bind to a DNA scaffold called the origin of replication (*oriC*). Orisomes perform the essential tasks of unwinding *oriC* and loading the replicative helicase, and orisome assembly is tightly regulated in the cell cycle to ensure chromosome replication begins only once. Only a few bacterial orisomes have been fully characterized, and while this lack of information complicates identification of all features that could be targeted, examination of assembly stages and orisome regulatory mechanisms may provide direction for some effective inhibitory strategies. In this perspective, we review current knowledge about orisome assembly and regulation, and identify potential targets that, when inhibited pharmacologically, would prevent bacterial chromosome replication.

**Keywords:** antibiotic discovery, orisome, *oriC*, DnaA, initiation of bacterial DNA replication

## INTRODUCTION

The increase in life-threatening infections caused by multi-drug resistant bacteria has caused an urgent need for new antibiotics. Prevalence of drug-resistant bacteria can be partly attributed to over-use of antibiotics, both clinically and agriculturally (Ventola, 2015), but antibiotic resistance is an ancient phenomenon (D'Costa et al., 2011), and selection of resistant organisms is a predictable and inevitable consequence of antibiotic use. Complicating the problem is lack of diversity in current antibiotic targets; of the approximately 200 essential genes identified in bacteria, only a handful are currently targeted (Lewis, 2013). Because recent drug discovery efforts have focused largely on modifying existing scaffolds, any new drug that acts on molecular targets in the few exploited processes risks encountering pre-selected, resistance-causing mutations (Barker, 1999). Therefore, one logical way to combat antibiotic resistance is to expand the set of targeted essential processes and proteins. One unexploited process is assembly of the orisome, the nucleoprotein complex that mediates initiation of bacterial chromosome replication, a critical event in the bacterial cell cycle (Leonard and Grimwade, 2015). In this perspective, we review orisome assembly, and address whether or not orisomes contain molecular targets that are not only novel, but which might also lead to the development of clinically useful antibiotics.



## ORISOME ASSEMBLY

All bacteria must duplicate their genomes before they divide into two identical daughter cells. With a few exceptions, all bacteria share fundamental molecular machinery responsible for triggering new rounds of DNA synthesis, comprising a unique chromosomal replication origin, *oriC*, and the conserved initiator protein, DnaA, a member of the AAA+ family of ATPases. The nucleoprotein complex formed by these two components is termed the orisome, which, when fully assembled, unwinds *oriC* DNA, and recruits replicative helicase, preparing the origin for the two replisomes required to bi-directionally replicate the circular genome (Wolański et al., 2014; Leonard and Grimwade, 2015).

The model for orisome assembly (**Figure 1**) is based largely on studies using *Escherichia coli* (Leonard and Grimwade, 2011, 2015). The orisome assembles from a persistent scaffold comprising three molecules of DnaA, interacting with three high affinity recognition boxes (R1, R2, and R4) (Cassler et al., 1995). The scaffold (stage 1) establishes a conformation of *oriC* that prevents premature unwinding and allows negative regulation by the DNA-bending protein Fis (Kaur et al., 2014). This scaffold also recruits and positions additional DnaA molecules for the next assembly stage (stage 2) (Miller et al., 2009). In stage 2, the N-terminal domain of DnaA bound to the high affinity R1 or R4 sites recruits DnaA to the proximal low affinity site (R5M or C1), followed by progressive binding of DnaA to the remaining lower affinity (non-consensus) binding sites; these sites preferentially bind DnaA-ATP (McGarry et al., 2004; Rozgaja et al., 2011). In the left region of *oriC*, DNA bending, assisted by the IHF protein, brings R1 and R5M into proximity to facilitate the cooperative DnaA site filling in *oriC*'s left half (Grimwade et al., 2000). Occupation of low affinity sites is required for the final stage (stage 3), when AT-rich DNA in a DNA Unwinding Element (DUE) is unwound, and DnaA-ATP associates with the single-stranded region (Yung and Kornberg, 1989; Speck and Messer, 2001), either in the form of a compact filament, or through interactions between ssDNA and domain III of DnaA bound to the left array of sites (Duderstadt et al., 2010; Ozaki and Katayama, 2012). DnaA in the DUE then recruits the replicative helicase and the helicase loader (DnaB and DnaC, respectively, in *E. coli*) (Sutton et al., 1998; Mott et al., 2008).

The instructions for orisome assembly are carried in all bacterial *oriC*'s in the form of precisely positioned recognition sites that direct DnaA binding (Rozgaja et al., 2011). DnaA is highly conserved and the consensus DnaA recognition motif in *E. coli* (5'-TTATCCACA) is also utilized by most bacteria (Schaper and Messer, 1995; Speck et al., 1997). However, there can be significant differences in the affinity each DnaA has for recognition sequences, particularly those that diverge from consensus (Zawilak-Pawlik et al., 2005; Ozaki et al., 2006). In addition, a database (DoriC<sup>1</sup>) (Gao et al., 2013) of over 1000 bacterial replication origins reveals a surprising variation in the arrangement, orientation and number of consensus or near consensus DnaA recognition sites among the *oriC*'s of different

bacterial types. Thus, although all orisomes contain a conserved protein (DnaA) and all perform the same essential function of origin activation, there is little obvious similarity in the set of instructions used to assemble them. How this diversity influences individual assembly stages and the transitions between those stages is not yet clear, and this lack of information could hamper identification of some conserved features essential for the mechanical aspects of origin activation that could be used as targets in antibiotic screens. Studies on orisomes outside of *E. coli* are ongoing, and the reader is referred to recent reviews discussing orisome assembly in different bacterial types (Wolański et al., 2014), as well as a review that includes strategies for rapid comparative analyses of diverse orisomes (Leonard and Grimwade, 2015).

## ORISOME REGULATORY MECHANISMS: A POTENTIAL GUIDE TO EFFECTIVE DRUG TARGETS?

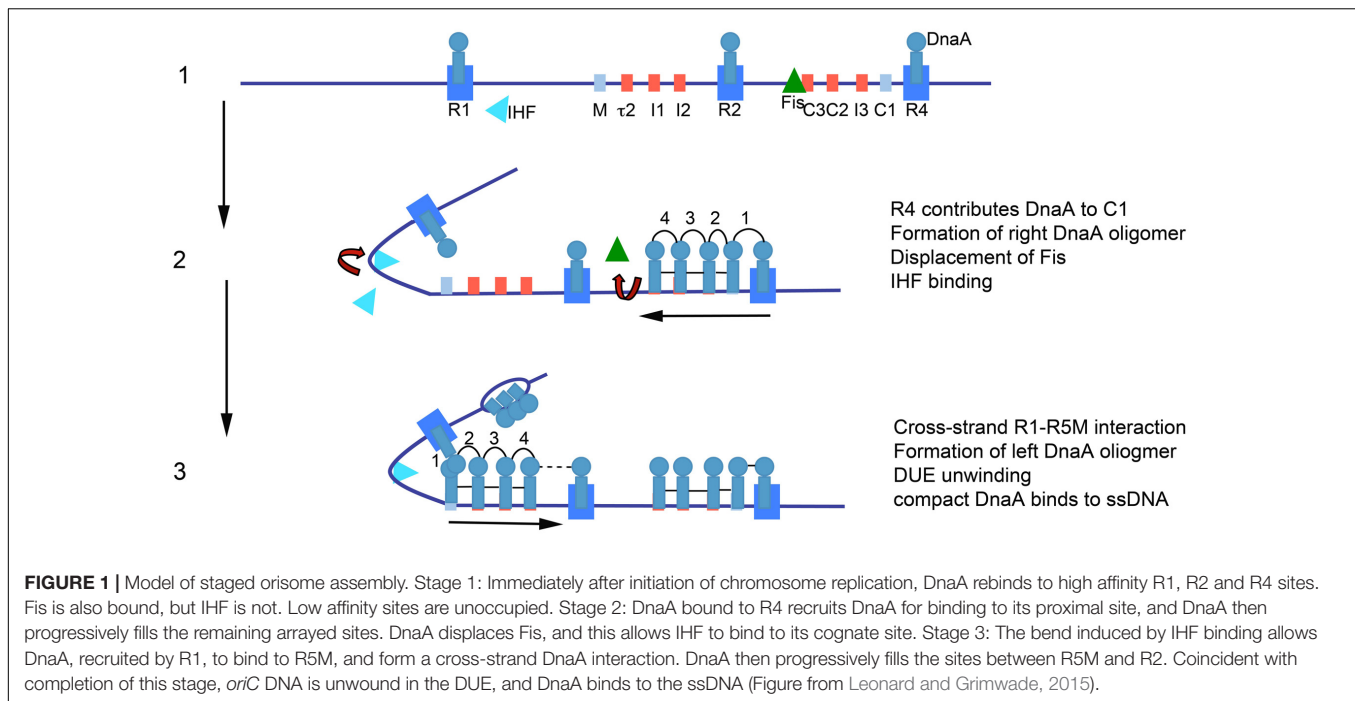
Because more research is required before there is a unified paradigm for how orisomes trigger initiation, the best current strategy for identifying orisome targets may be to examine molecular mechanisms that regulate assembly. Logically, conserved mechanisms that inhibit orisome assembly will prevent initiation, and should provide "proof of principle" to justify targets as appropriate for pharmacological inhibition.

All orisomes are tightly regulated so that they trigger initiation of chromosome replication once, only once, and at the correct time in the cell cycle (Skarstad and Katayama, 2013). Delayed, or under-initiation leads to eventual chromosome loss, while re-initiation from the same origin can result in replication fork collapse and genome instability (Simmons et al., 2004). Like orisome assembly, regulation is best understood in *E. coli*, where two non-competing mechanisms, regulation of DnaA/*oriC* interactions, and regulation of cellular DnaA-ATP levels, predominate. Below, we review these two mechanisms and evaluate their possible utility as drug targets.

### Orisome Regulation by Controlling DnaA-*oriC* Interactions

In *E. coli*, DnaA binding to *oriC* is controlled both before and immediately after initiation by mechanisms that prevent completion of orisome assembly stages 2 and 3 (Leonard and Grimwade, 2005). Before initiation, the DNA bending protein Fis helps maintain the origin in a conformation that reduces DnaA's ability to bind low affinity sites, until levels of DnaA increase enough to displace Fis from its recognition site (Ryan et al., 2004; Kaur et al., 2014). Since *E. coli oriC* contains multiple low affinity DnaA binding sites that preferentially bind DnaA-ATP (McGarry et al., 2004; Kawakami et al., 2005), orisome assembly cannot be completed until DnaA-ATP levels rise to a critical level. (Regulation of DnaA-ATP levels is discussed below.) After initiation, the SeqA protein binds hemimethylated GATC motifs in *oriC*, several of which are inside or overlap low affinity DnaA recognition sites (Lu et al., 1994; Skarstad et al., 2001). SeqA

<sup>1</sup><http://tubic.tju.edu.cn/doric/index.php>



blocks DnaA-ATP from re-occupying low affinity sites and the DUE region for approximately one third of the cell cycle (Nievera et al., 2006).

It isn't known how many bacterial origins contain low affinity recognition sites with preference for DnaA-ATP, and not all bacteria use Fis or SeqA to regulate orisome assembly (Brézellec et al., 2006; Madiraju et al., 2006). Regardless, the basic paradigm of controlling DnaA's access to *oriC* as a way of regulating orisome assembly can be found in many bacterial types. For example, response regulators CtrA, MtrA, and HP1021 inhibit DnaA occupation of *oriC* in *Caulobacter crescentus*, *Mycobacterium tuberculosis*, and *Helicobacter pylori*, respectively, and by doing so, help prevent untimely initiations (Taylor et al., 2011; Donczew et al., 2015; Purushotham et al., 2015). *H. pylori* also uses DNA topology to regulate DnaA/*oriC* interactions (Donczew et al., 2014). In *Bacillus subtilis*, several proteins have been identified that negatively regulate initiation by inhibiting cooperative binding of DnaA at *oriC*; these include YabA (Merrikh and Grossman, 2011; Scholefield and Murray, 2013), DnaD (Bonilla and Grossman, 2012; Scholefield and Murray, 2013), and Soj (Scholefield et al., 2012). In several systems, orisome assembly is also controlled by positive regulators that increase DnaA binding to low affinity sites. In *E. coli* and *Caulobacter crescentus*, low affinity DnaA binding is stimulated by the DNA bending protein IHF (Grimwade et al., 2000; Siam et al., 2003). Additionally, the *E. coli* DiaA protein (Ishida et al., 2004), and its homolog in *H. pylori*, HobA (Natrajan et al., 2007; Zawilak-Pawlik et al., 2007, 2011), bind to DnaA's domain I and increase weak site occupation.

The studies described above suggest that several different regions of DnaA could be targeted to inhibit DnaA binding. Obviously, blocking the DNA binding domain (domain IV)

should inhibit all stages of orisome formation. Although protein–DNA interactions have not traditionally been considered to be “druggable” targets, recent studies have reported success in identifying inhibitors of DNA binding (Huang et al., 2016; Grimley et al., 2017). Further, inhibition of the self-oligomerization regions of DnaA in domains I and III should prevent cooperative binding and thus assembly stages 2 and 3 (Kawakami et al., 2005; Miller et al., 2009; Duderstadt et al., 2010; Scholefield and Murray, 2013). Like protein–DNA interactions, protein–protein interactions have not traditionally been favored as drug targets, but recent reports raise optimism that targeting DnaA oligomerization could be successful (Marceau et al., 2013; Voter et al., 2017).

Several other must be resolved before inhibition of DnaA's access to *oriC* can be determined to be a practical antimicrobial strategy. First, it is not yet clear how much binding must be prevented to inhibit replication. All origins contain multiple DnaA binding sites (Leonard and Mechali, 2013), and studies that removed or inactivated DnaA recognition sites in *E. coli* chromosomal *oriC* revealed a tremendous plasticity in orisome assembly. Remarkably, deletion of the entire right region of *oriC* is tolerated in slow growing cells (Stepankiw et al., 2009). Additionally, directed mutations that knocked out binding to individual chromosomal *oriC* sites had little effect on viability (Weigel et al., 2001; Riber et al., 2009; Kaur et al., 2014). However, eliminating binding to more than one high affinity site did cause loss of viability (Kaur et al., 2014). Similar plasticity was noted in SeqA regulation of the number of occupied DnaA sites; even though loss of SeqA binding would be expected to allow DnaA re-binding at some *oriC* sites after initiation, mutating individual GATCs had little effect on initiation synchrony (Jha and Chattoraj, 2016). In *Bacillus*, some individual chromosomal

*oriC* DnaA binding sites were shown to be essential, but others were not (Richardson et al., 2016). These studies, although by no means comprehensive, suggest that any pharmacological strategy should aim to inhibit DnaA binding at a majority of *oriC* sites, at least until future orisome studies reveal which sites are needed to assemble sub-complexes that carry out the essential mechanical functions of origin activation. Additionally, several studies suggest that assays used to screen for inhibitors of DnaA binding should be based on inhibiting chromosomal *oriC* rather than cloned origins, since inactivating individual sites is much more detrimental to plasmid *oriC* function (Weigel et al., 2001). Also, given the diversity in bacterial origin configurations (Leonard and Mechali, 2013), screens using a single bacterial type might not be sufficient to identify agents that act against a broad spectrum of bacteria. It might be necessary to utilize multiple types of bacteria, unless methodology is developed that allows the function of any chromosomal origin to be examined in an easily cultured strain. One strategy, involving heterologous origin transplantation, was described in a recent review (Leonard and Grimwade, 2015).

## Orisome Regulation by Controlling DnaA-ATP Levels

Based on seminal studies of *in vitro* *E. coli* DNA replication by the Kornberg lab (Sekimizu et al., 1987), DnaA-ATP is the active initiator form, and it is widely accepted that all bacteria share the requirement for DnaA-ATP in origin activation. In *E. coli*, DnaA-ATP levels are tightly regulated during the cell cycle to ensure precise initiation timing. Prior to the initiation step, DnaA-ATP levels rise due to new synthesis and a combination of recharging systems that include the DARS loci and acidic phospholipids in the membrane, reviewed in Skarstad and Katayama (2013). After initiation, the synthesis of DnaA-ATP is repressed for 1/3 of the cell cycle by SeqA, which binds to hemi-methylated GATC motifs in the *dnaA* promoter (Campbell and Kleckner, 1990). To inactivate DnaA-ATP, DnaA's intrinsic ATPase activity is stimulated by the Hda protein associated with the  $\beta$ -clamp (DnaN) (Su'etsugu et al., 2004; Kim et al., 2017). Excess DnaA-ATP can also bind to a high capacity locus, termed *datA* (Kitagawa et al., 1998), which also stimulates DnaA-ATP hydrolysis (Kasho and Katayama, 2013).

The critical importance of mechanisms regulating DnaA-ATP levels in *E. coli* is demonstrated by the lethality observed in mutants, such as *dnaA(cos)* and *hda* null, that have lost the ability to inactivate DnaA-ATP by hydrolysis (Nishida et al., 2002; Felczak and Kaguni, 2009). *DnaA(cos)* carries two amino acid substitutions, one that prevents nucleotide binding (A184V), and another that stabilizes the mutated form (Y271H) (Simmons and Kaguni, 2003). Cells harboring *dnaA(cos)* are non-viable at 30°C, most likely due to over-initiation that results in co-directional replication fork collisions at stalled forks, leading to catastrophic double-stranded breaks (Simmons et al., 2004). A similar lethal phenotype is seen when Hda is inactivated, unless suppressor mutations arise (Riber et al., 2009; Charbon et al., 2011). Interestingly, although diverse suppressor mutations have been identified (Charbon et al., 2011), they all seem to cause tolerance of over-initiation by decreasing the chance of

fork collisions, either by reducing initiation frequencies, or by preventing DNA lesions, such as oxidative DNA damage, that would slow forks (Charbon et al., 2014, 2017).

There are several aspects of DnaA inactivation mutants that are relevant to identifying antibiotic targets. First, lethality is caused by increasing, rather than decreasing the initiation frequency (Simmons et al., 2004). The run-away replication observed in *DnaA(cos)* mutants correlates with the inability to bind adenine nucleotide (Simmons et al., 2003), although it is not clear why loss of nucleotide binding leads to over-replication rather than orisome inactivation. Second, it is not yet known how many other bacterial types use regulation of DnaA-ATP levels as a regulatory mechanism. While some bacteria, such as *Caulobacter* and most enterobacteria, appear to have homologs of *hda* (Wargachuk and Marczynski, 2015), others, such as *Bacillus*, *Staphylococcus*, and *H. pylori*, do not (Katayama et al., 2010). DnaA in *B. subtilis* and *S. aureus* exchange bound ADP for ATP much more rapidly than *E. coli* does (Kurokawa et al., 2009; Bonilla and Grossman, 2012), and negative regulation of orisomes in these bacteria is focused mainly on DnaA-DNA interactions. Thus, screens to identify stimulators of DnaA hydrolysis may be ineffective in identifying broad-spectrum antimicrobials. In contrast, the growth inhibition/lethality caused when DnaA can't hydrolyze ATP suggests that identification of inhibitors of ATP binding or ATPase activity, causing lethality by over-initiation, may be more fruitful. While targeting of the ATPase of AAA+ proteins is still in its infancy there are reports of successful inhibition of this protein class (Chou et al., 2011; Firestone et al., 2012). Targeting of DnaA's ATPase, however, could generate suppressor mutations that reduce fork collisions (Charbon et al., 2017) and thus be prone to rapid resistance development. Possibly, this could be resolved by combination with an agent that inhibits DNA repair to counteract the actions of suppressor mutations (Simmons et al., 2004; Suter and Lovett, 2006).

## ADDITIONAL CONSIDERATIONS IN TARGETING ORISOME FUNCTION

Obviously, any antibiotic acting on the orisome must enter the bacterial cell. This presents a problem with all bacteria, but particularly Gram negative bacteria, where the relatively impermeable outer membrane presents a potential barrier to drug delivery (Lewis, 2013; Brown, 2016). Until more is known about transport across the outer membrane, successful platforms to discover drugs affecting orisomes or any other intracellular target are likely to require living cells to augment or replace *in vitro* biochemical assays. While screen development is beyond the scope of this Perspective, we note that one cell-based assay, to identify agents that allow *dnaA(cos)* cells to grow at non-permissive temperature, has been described (Fossum et al., 2008), but failed to identify any small molecule inhibitors of DnaA function in a limited trial screen, although it is possible that lead compounds could be identified by screening a much larger library.



Of greater concern is generation of intra- or extra-genic suppressors, particularly if a new drug causes over-replication. Unfortunately, bacteria are adept in their ability to survive initiation perturbation. In cases where rapid development of resistance is expected, hybrid antibiotics or combination chemotherapy, where orisome inhibitors are combined with drugs that act on different pathways, should be considered. Alternatively, it might be useful to target features within DnaA that are shared by other proteins, since the majority of currently used successful antibiotics delay resistance development by attacking more than one target (Silver, 2011; Brown and Wright, 2016). One possible shared motif is the AAA+ domain, since the AAA+ domain of DnaC is quite similar to that of DnaA (Mott et al., 2008). Interestingly, hydrolysis of the ATP bound to DnaC is required before DnaB helicase can function (Mott et al., 2008), and it may be possible to identify inhibitors of DnaA's intrinsic ATPase that also inhibit DnaB activation.

It is interesting that no natural product that inhibits orisome function has been identified in many years of antibiotic screening. This may be because the assays are not designed to identify drugs

inhibiting this essential process, or that targeting the orisome is an inherently risky competition strategy for any bacteria, and so it rarely evolves. Regardless, the orisome appears to have potential as a novel and effective drug target, and its usefulness in antibiotic discovery should increase as more studies reveal conserved and non-conserved features of orisome assembly among bacterial types.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENT

The work in our laboratories was supported by Public Health Service grant GM54042. Publication of this article was funded in part by the Open Access Subvention Fund and the Florida Tech Libraries.

## REFERENCES

- Barker, K. F. (1999). Antibiotic resistance: a current perspective. *Br. J. Clin. Pharmacol.* 48, 109–124. doi: 10.1046/j.1365-2125.1999.00997.x
- Bonilla, C. Y., and Grossman, A. D. (2012). The primosomal protein DnaD inhibits cooperative DNA binding by the replication initiator DnaA in *Bacillus subtilis*. *J. Bacteriol.* 194, 5110–5117. doi: 10.1128/JB.00958-12
- Brézellec, P., Hoebeke, M., Hiet, M. S., Pasek, S., and Ferat, J. L. (2006). DomainSieve: a protein domain-based screen that led to the identification of dam-associated genes with potential link to DNA maintenance. *Bioinformatics* 22, 1935–1941. doi: 10.1093/bioinformatics/btl336
- Brown, D. G. (2016). Drug discovery strategies to outer membrane targets in Gram-negative pathogens. *Bioorg. Med. Chem.* 24, 6320–6331. doi: 10.1016/j.bmc.2016.05.004
- Brown, E. D., and Wright, G. D. (2016). Antibacterial drug discovery in the resistance era. *Nature* 529, 336–343. doi: 10.1038/nature17042
- Campbell, J. L., and Kleckner, N. (1990). *E. coli* oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* 62, 967–979. doi: 10.1016/0092-8674(90)90271-F
- Cassler, M. R., Grimwade, J. E., and Leonard, A. C. (1995). Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *EMBO J.* 14, 5833–5841.
- Charbon, G., Bjørn, L., Mendoza-Chamizo, B., Frimodt-Møller, J., and Løbner-Olesen, A. (2014). Oxidative DNA damage is instrumental in hyperreplication stress-induced inviability of *Escherichia coli*. *Nucleic Acids Res.* 42, 13228–13241. doi: 10.1093/nar/gku1149
- Charbon, G., Riber, L., Cohen, M., Skovgaard, O., Fujimitsu, K., Katayama, T., et al. (2011). Suppressors of DnaA(ATP) imposed overinitiation in *Escherichia coli*. *Mol. Microbiol.* 79, 914–928. doi: 10.1111/j.1365-2958.2010.07493.x
- Charbon, G., Riber, L., and Løbner-Olesen, A. (2017). Countermeasures to survive excessive chromosome replication in *Escherichia coli*. *Curr. Genet.* doi: 10.1007/s00294-017-0725-4 [Epub ahead of print].
- Chou, T. F., Brown, S. J., Minond, D., Nordin, B. E., Li, K., Jones, A. C., et al. (2011). Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4834–4839. doi: 10.1073/pnas.1015312108
- D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W., Schwarz, C., et al. (2011). Antibiotic resistance is ancient. *Nature* 477, 457–461. doi: 10.1038/nature10388
- Donczew, R., Makowski, L., Jaworski, P., Bezulka, M., Nowaczyk, M., Zakrzewska-Czerwińska, J., et al. (2015). The atypical response regulator HP1021 controls formation of the *Helicobacter pylori* replication initiation complex. *Mol. Microbiol.* 95, 297–312. doi: 10.1111/mmi.12866
- Donczew, R., Mielke, T., Jaworski, P., Zakrzewska-Czerwińska, J., and Zawilak-Pawlik, A. (2014). Assembly of *Helicobacter pylori* initiation complex is determined by sequence-specific and topology-sensitive DnaA-oriC interactions. *J. Mol. Biol.* 426, 2769–2782. doi: 10.1016/j.jmb.2014.05.018
- Duderstadt, K. E., Mott, M. L., Crisona, N. J., Chuang, K., Yang, H., and Berger, J. M. (2010). Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.* 285, 28229–28239. doi: 10.1074/jbc.M110.147975
- Felczak, M. M., and Kaguni, J. M. (2009). DnaAcos hyperinitiates by circumventing regulatory pathways that control the frequency of initiation in *Escherichia coli*. *Mol. Microbiol.* 72, 1348–1363. doi: 10.1111/j.1365-2958.2009.06724.x
- Firestone, A. J., Weinger, J. S., Maldonado, M., Barlan, K., Langston, L. D., O'Donnell, M., et al. (2012). Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein. *Nature* 484, 125–129. doi: 10.1038/nature10936
- Fossum, S., De Pascale, G., Weigel, C., Messer, W., Donadio, S., and Skarstad, K. (2008). A robust screen for novel antibiotics: specific knockout of the initiator of bacterial DNA replication. *FEMS Microbiol. Lett.* 281, 210–214. doi: 10.1111/j.1574-6968.2008.01103.x
- Gao, F., Luo, H., and Zhang, C. T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990
- Grimley, E., Liao, C., Ranghini, E. J., Nikolovska-Coleska, Z., and Dressler, G. R. (2017). Inhibition of Pax2 transcription activation with a small molecule that targets the DNA binding domain. *ACS Chem. Biol.* 12, 724–734. doi: 10.1021/acschembio.6b00782
- Grimwade, J. E., Ryan, V. T., and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled oriC of *Escherichia coli*. *Mol. Microbiol.* 35, 835–844. doi: 10.1046/j.1365-2958.2000.01755.x
- Huang, W., Dong, Z., Chen, Y., Wang, F., Wang, C. J., Peng, H., et al. (2016). Small-molecule inhibitors targeting the DNA-binding domain of STAT3 suppress tumor growth, metastasis and STAT3 target gene expression in vivo. *Oncogene* 35, 783–792. doi: 10.1038/ncr.2015.215
- Ishida, T., Akimitsu, N., Kashioka, T., Hatano, M., Kubota, T., Ogata, Y., et al. (2004). DiaA, a novel DnaA-binding protein, ensures the timely initiation of *Escherichia coli* chromosome replication. *J. Biol. Chem.* 279, 45546–45555. doi: 10.1074/jbc.M402762200



- Jha, J. K., and Chattoraj, D. K. (2016). Inactivation of individual SeqA binding sites of the *E. coli* origin reveals robustness of replication initiation synchrony. *PLOS ONE* 11:e0166722. doi: 10.1371/journal.pone.0166722
- Kasho, K., and Katayama, T. (2013). DnaA binding locus *datA* promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 936–941. doi: 10.1073/pnas.1212070110
- Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nat. Rev. Microbiol.* 8, 163–170. doi: 10.1038/nrmicro2314
- Kaur, G., Vora, M. P., Czerwonka, C. A., Rozgaja, T. A., Grimwade, J. E., and Leonard, A. C. (2014). Building the bacterial orisome: high-affinity DnaA recognition plays a role in setting the conformation of *oriC* DNA. *Mol. Microbiol.* 91, 1148–1163. doi: 10.1111/mmi.12525
- Kawakami, H., Keyamura, K., and Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J. Biol. Chem.* 280, 27420–27430. doi: 10.1074/jbc.M502764200
- Kim, J. S., Nanfara, M. T., Chodavarapu, S., Jin, K. S., Babu, V. M. P., Ghazy, M. A., et al. (2017). Dynamic assembly of Hda and the sliding clamp in the regulation of replication licensing. *Nucleic Acids Res.* 45, 3888–3905. doi: 10.1093/nar/gkx081
- Kitagawa, R., Ozaki, T., Moriya, S., and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev.* 12, 3032–3043. doi: 10.1101/gad.12.19.3032
- Kurokawa, K., Mizumura, H., Takaki, T., Ishii, Y., Ichihashi, N., Lee, B. L., et al. (2009). Rapid exchange of bound ADP on the *Staphylococcus aureus* replication initiation protein DnaA. *J. Biol. Chem.* 284, 34201–34210. doi: 10.1074/jbc.M109.060681
- Leonard, A. C., and Grimwade, J. E. (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Mol. Microbiol.* 55, 978–985. doi: 10.1111/j.1365-2958.2004.04467.x
- Leonard, A. C., and Grimwade, J. E. (2011). Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.* 65, 19–35. doi: 10.1146/annurev-micro-090110-102934
- Leonard, A. C., and Grimwade, J. E. (2015). The orisome: structure and function. *Front. Microbiol.* 6:545. doi: 10.3389/fmicb.2015.00545
- Leonard, A. C., and Mechali, M. (2013). DNA replication origins. *Cold Spring Harb. Perspect. Biol.* 5:a010116. doi: 10.1101/cshperspect.a010116
- Lewis, K. (2013). Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* 12, 371–387. doi: 10.1038/nrd3975
- Lu, M., Campbell, J. L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426. doi: 10.1016/0092-8674(94)90156-2
- Madiraju, M. V., Moomey, M., Neuenschwander, P. F., Muniruzzaman, S., Yamamoto, K., Grimwade, J. E., et al. (2006). The intrinsic ATPase activity of *Mycobacterium tuberculosis* DnaA promotes rapid oligomerization of DnaA on *oriC*. *Mol. Microbiol.* 59, 1876–1890. doi: 10.1111/j.1365-2958.2006.05068.x
- Marceau, A. H., Bernstein, D. A., Walsh, B. W., Shapiro, W., Simmons, L. A., and Keck, J. L. (2013). Protein interactions in genome maintenance as novel antibacterial targets. *PLOS ONE* 8:e58765. doi: 10.1371/journal.pone.0058765
- McGarry, K. C., Ryan, V. T., Grimwade, J. E., and Leonard, A. C. (2004). Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2811–2816. doi: 10.1073/pnas.0400340101
- Merrikh, H., and Grossman, A. D. (2011). Control of the replication initiator DnaA by an anti-cooperativity factor. *Mol. Microbiol.* 82, 434–446. doi: 10.1111/j.1365-2958.2011.07821.x
- Miller, D. T., Grimwade, J. E., Betteridge, T., Rozgaja, T., Torgue, J. J., and Leonard, A. C. (2009). Bacterial origin recognition complexes direct assembly of higher-order DnaA oligomeric structures. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18479–18484. doi: 10.1073/pnas.0909472106
- Mott, M. L., Erzberger, J. P., Coons, M. M., and Berger, J. M. (2008). Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* 135, 623–634. doi: 10.1016/j.cell.2008.09.058
- Natrajan, G., Hall, D. R., Thompson, A. C., Gutsche, I., and Terradot, L. (2007). Structural similarity between the DnaA-binding proteins HobA (HP1230) from *Helicobacter pylori* and DiaA from *Escherichia coli*. *Mol. Microbiol.* 65, 995–1005. doi: 10.1111/j.1365-2958.2007.05843.x
- Nievera, C., Tor, J. J., Grimwade, J. E., and Leonard, A. C. (2006). SeqA blocking of DnaA-*oriC* interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell* 24, 581–592. doi: 10.1016/j.molcel.2006.09.016
- Nishida, S., Fujimitsu, K., Sekimizu, K., Ohmura, T., Ueda, T., and Katayama, T. (2002). A nucleotide switch in the *Escherichia coli* DnaA protein initiates chromosomal replication: evidence from a mutant DnaA protein defective in regulatory ATP hydrolysis in vitro and in vivo. *J. Biol. Chem.* 277, 14986–14995. doi: 10.1074/jbc.M108303200
- Ozaki, S., Fujimitsu, K., Kurumizaka, H., and Katayama, T. (2006). The DnaA homolog of the hyperthermophilic eubacterium *Thermotoga maritima* forms an open complex with a minimal 149-bp origin region in an ATP-dependent manner. *Genes Cells* 11, 425–438. doi: 10.1111/j.1365-2443.2006.00950.x
- Ozaki, S., and Katayama, T. (2012). Highly organized DnaA-*oriC* complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832
- Purushotham, G., Sarva, K. B., Blaszczyk, E., Rajagopalan, M., and Madiraju, M. V. (2015). *Mycobacterium tuberculosis* *oriC* sequestration by MtrA response regulator. *Mol. Microbiol.* 98, 586–604. doi: 10.1111/mmi.13144
- Riber, L., Fujimitsu, K., Katayama, T., and Løbner-Olesen, A. (2009). Loss of Hda activity stimulates replication initiation from I-box, but not R4 mutant origins in *Escherichia coli*. *Mol. Microbiol.* 71, 107–122. doi: 10.1111/j.1365-2958.2008.06516.x
- Richardson, T. T., Harran, O., and Murray, H. (2016). The bacterial DnaA-trio replication origin element specifies single-stranded DNA initiator binding. *Nature* 534, 412–416. doi: 10.1038/nature17962
- Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A. C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in *oriC* guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly. *Mol. Microbiol.* 82, 475–488. doi: 10.1111/j.1365-2958.2011.07827.x
- Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E., and Leonard, A. C. (2004). *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol. Microbiol.* 51, 1347–1359. doi: 10.1046/j.1365-2958.2003.03906.x
- Schaper, S., and Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* 270, 17622–17626. doi: 10.1074/jbc.270.29.17622
- Scholefield, G., Errington, J., and Murray, H. (2012). Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J.* 31, 1542–1555. doi: 10.1038/emboj.2012.6
- Scholefield, G., and Murray, H. (2013). YabA and DnaD inhibit helix assembly of the DNA replication initiation protein DnaA. *Mol. Microbiol.* 90, 147–159. doi: 10.1111/mmi.12353
- Sekimizu, K., Bramhill, D., and Kornberg, A. (1987). ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* 50, 259–265. doi: 10.1016/0092-8674(87)90221-2
- Siam, R., Brassinga, A. K., and Marczyński, G. T. (2003). A dual binding site for integration host factor and the response regulator CtrA inside the *Caulobacter crescentus* replication origin. *J. Bacteriol.* 185, 5563–5572. doi: 10.1128/JB.185.18.5563-5572.2003
- Silver, L. L. (2011). Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* 24, 71–109. doi: 10.1128/CMR.00030-10
- Simmons, L. A., Breier, A. M., Cozzarelli, N. R., and Kaguni, J. M. (2004). Hyperinitiation of DNA replication in *Escherichia coli* leads to replication fork collapse and inviability. *Mol. Microbiol.* 51, 349–358. doi: 10.1046/j.1365-2958.2003.03842.x
- Simmons, L. A., Felczak, M., and Kaguni, J. M. (2003). DnaA Protein of *Escherichia coli*: oligomerization at the *E. coli* chromosomal origin is required for initiation and involves specific N-terminal amino acids. *Mol. Microbiol.* 49, 849–858. doi: 10.1046/j.1365-2958.2003.03603.x
- Simmons, L. A., and Kaguni, J. M. (2003). The DnaAcos allele of *Escherichia coli*: hyperactive initiation is caused by substitution of A184V and Y271H, resulting in defective ATP binding and aberrant DNA replication control. *Mol. Microbiol.* 47, 755–765. doi: 10.1046/j.1365-2958.2003.03333.x

- Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5:a012922. doi: 10.1101/cshperspect.a012922
- Skarstad, K., Torheim, N., Wold, S., Lurz, R., Messer, W., Fossum, S., et al. (2001). The *Escherichia coli* SeqA protein binds specifically to two sites in fully and hemimethylated oriC and has the capacity to inhibit DNA replication and affect chromosome topology. *Biochimie* 83, 49–51. doi: 10.1016/S0300-9084(00)01207-4
- Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* 20, 1469–1476. doi: 10.1093/emboj/20.6.1469
- Speck, C., Weigel, C., and Messer, W. (1997). From footprint to toeprint: a close-up of the DnaA box, the binding site for the bacterial initiator protein DnaA. *Nucleic Acids Res.* 25, 3242–3247. doi: 10.1093/nar/25.16.3242
- Stepankiw, N., Kaidow, A., Boye, E., and Bates, D. (2009). The right half of the *Escherichia coli* replication origin is not essential for viability, but facilitates multi-forked replication. *Mol. Microbiol.* 74, 467–479. doi: 10.1111/j.1365-2958.2009.06877.x
- Su'tsugu, M., Takata, M., Kubota, T., Matsuda, Y., and Katayama, T. (2004). Molecular mechanism of DNA replication-coupled inactivation of the initiator protein in *Escherichia coli*: interaction of DnaA with the sliding clamp-loaded DNA and the sliding clamp-Hda complex. *Genes Cells* 9, 509–522. doi: 10.1111/j.1356-9597.2004.00741.x
- Sutera, V. A., and Lovett, S. T. (2006). The role of replication initiation control in promoting survival of replication fork damage. *Mol. Microbiol.* 60, 229–239. doi: 10.1111/j.1365-2958.2006.05093.x
- Sutton, M. D., Carr, K. M., Vicente, M., and Kaguni, J. M. (1998). *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J. Biol. Chem.* 273, 34255–34262. doi: 10.1074/jbc.273.51.34255
- Taylor, J. A., Ouimet, M. C., Wargachuk, R., and Marczyński, G. T. (2011). The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol. Microbiol.* 82, 312–326. doi: 10.1111/j.1365-2958.2011.07785.x
- Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P T* 40, 277–283.
- Voter, A. F., Killoran, M. P., Ananiev, G. E., Wildman, S. A., Hoffmann, F. M., and Keck, J. L. (2017). A high-throughput screening strategy to identify inhibitors of SSB protein-protein interactions in an academic screening facility. *SLAS Discov.* doi: 10.1177/2472555217712001 [Epub ahead of print].
- Wargachuk, R., and Marczyński, G. T. (2015). The *Caulobacter crescentus* homolog of DnaA (HdaA) also regulates the proteolysis of the replication initiator protein DnaA. *J. Bacteriol.* 197, 3521–3532. doi: 10.1128/JB.00460-15
- Weigel, C., Messer, W., Preiss, S., Welzck, M., Morigen, and Boye, E. (2001). The sequence requirements for a functional *Escherichia coli* replication origin are different for the chromosome and a minichromosome. *Mol. Microbiol.* 40, 498–507. doi: 10.1046/j.1365-2958.2001.02409.x
- Wolański, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwińska, J. (2014). oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 5:735. doi: 10.3389/fmicb.2014.00735
- Yung, B. Y., and Kornberg, A. (1989). The dnaA initiator protein binds separate domains in the replication origin of *Escherichia coli*. *J. Biol. Chem.* 264, 6146–6150.
- Zawilak-Pawlik, A., Donczew, R., Szafrński, S., Mackiewicz, P., Terradot, L., and Zakrzewska-Czerwińska, J. (2011). DiaA/HobA and DnaA: a pair of proteins co-evolved to cooperate during bacterial oriosome assembly. *J. Mol. Biol.* 408, 238–251. doi: 10.1016/j.jmb.2011.02.045
- Zawilak-Pawlik, A., Kois, A., Majka, J., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Messer, W., et al. (2005). Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem. J.* 389, 471–481. doi: 10.1042/BJ20050143
- Zawilak-Pawlik, A., Kois, A., Stingl, K., Boneca, I. G., Skrobuk, P., Piotr, J., et al. (2007). HobA—a novel protein involved in initiation of chromosomal replication in *Helicobacter pylori*. *Mol. Microbiol.* 65, 979–994. doi: 10.1111/j.1365-2958.2007.05853.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Grimwade and Leonard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Conjugation Inhibitors and Their Potential Use to Prevent Dissemination of Antibiotic Resistance Genes in Bacteria

Elena Cabezón\*, Fernando de la Cruz and Ignacio Arechaga\*

*Instituto de Biomedicina y Biotecnología de Cantabria (IBBTec), CSIC-Universidad de Cantabria and Departamento de Biología Molecular, Universidad de Cantabria, Santander, Spain*

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Universiti Sultan Zainal Abidin,  
Malaysia

### Reviewed by:

Charles Martin Dozois,  
Institut National de la Recherche  
Scientifique (INRS), Canada  
Irene Wagner-Dobler,  
Helmholtz-Zentrum für  
Infektionsforschung (HZI), Germany  
Radosław Pluta,  
International Institute of Molecular  
and Cell Biology in Warsaw (IIMCB),  
Poland

### \*Correspondence:

Elena Cabezón  
cabezona@unican.es  
Ignacio Arechaga  
arechagai@unican.es

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 04 August 2017

**Accepted:** 13 November 2017

**Published:** 30 November 2017

### Citation:

Cabezón E, de la Cruz F and  
Arechaga I (2017) Conjugation  
Inhibitors and Their Potential Use  
to Prevent Dissemination of Antibiotic  
Resistance Genes in Bacteria.  
Front. Microbiol. 8:2329.  
doi: 10.3389/fmicb.2017.02329

Antibiotic resistance has become one of the most challenging problems in health care. Bacteria conjugation is one of the main mechanisms whereby bacteria become resistant to antibiotics. Therefore, the search for specific conjugation inhibitors (COINs) is of interest in the fight against the spread of antibiotic resistances in a variety of laboratory and natural environments. Several compounds, discovered as COINs, are promising candidates in the fight against plasmid dissemination. In this review, we survey the effectiveness and toxicity of the most relevant compounds. Particular emphasis has been placed on unsaturated fatty acid derivatives, as they have been shown to be efficient in preventing plasmid invasiveness in bacterial populations. Biochemical and structural studies have provided insights concerning their potential molecular targets and inhibitory mechanisms. These findings open a new avenue in the search of new and more effective synthetic inhibitors. In this pursuit, the use of structure-based drug design methods will be of great importance for the screening of ligands and binding sites of putative targets.

**Keywords:** antibiotic resistance, type IV secretion systems, inhibitors, fatty acids, bacterial conjugation

## INTRODUCTION

Antibiotic resistance is becoming a major threat for human health (WHO, 2014). Widespread abuse of antibiotics in human health and food production are threatening our current health systems and challenge future care (Giske et al., 2008; Boucher et al., 2009; Rice, 2009; Chang et al., 2015). However, despite these threats, few new antibiotics are becoming available to fight against multi-resistant bugs (Gould and Bal, 2013; Hede, 2014). Bacterial conjugation is one of the main mechanisms whereby bacteria become resistant to antibiotics (Mazel and Davies, 1999; Waters, 1999). Thus, the search for specific conjugation inhibitors (COINs) is a foremost concern in the fight against the spread of antibiotic resistance genes (Smith and Romesberg, 2007; Baquero et al., 2011; Baym et al., 2016). In this pursuit, several compounds were reported to inhibit bacterial conjugation specifically, although most turned out to be unspecific growth inhibitors (Michel-Briand and Laporte, 1985; Hooper et al., 1989; Conter et al., 2002; Lujan et al., 2007; Nash et al., 2012).

Bacterial conjugation is a mechanism by which DNA is transferred between two bacterial cells. The process consists of two steps. In a first stage, DNA is mobilized by a set of proteins, encoded

by MOB genes. In a second step, DNA is transported across a secretion system [type IV secretion system (T4SS)] (de la Cruz et al., 2010; Cabezón et al., 2015). T4SS is a complex formed by proteins encoded by another set of genes, also known as MPF (mating pore formation) genes (Fernandez-Lopez et al., 2006). Conjugative Gram negative bacteria usually contain three MOB genes encoding proteins involved in DNA processing (Smillie et al., 2010). The most ubiquitous of these genes encodes a relaxase protein, which cleaves one of the plasmid strands at the origin of transfer (*oriT*) (Garcillán-Barcia et al., 2009) (Supplementary Figure S1A). Upon this, the relaxase protein remains covalently bound to the DNA at the 5' end. This nucleoprotein complex is recruited at the secretion channel (T4SS) with the assistance of the coupling protein, an ATPase present in most conjugative plasmids (Llosa et al., 2002; Tato et al., 2005) (Supplementary Figure S1B). A third, accessory protein is usually involved to ensure the correct DNA folding for the relaxase action (Moncalian and de la Cruz, 2004).

Conjugative T4SS are large multi-subunit complexes involved in substrate transport and pilus biogenesis. The simplest T4SS consists of 11 proteins, named VirB1 to VirB11, after *Agrobacterium tumefaciens* T4SS (Christie et al., 2005, 2014). This macromolecular complex spans across the inner and outer membranes and the periplasm in between. T4SS architecture is well-preserved in most conjugative bacteria, consisting of four distinct sections: the pilus, the core channel complex, the inner membrane platform and the hexameric ATPases that provide the energy for substrate transport and pilus biogenesis (Cabezón et al., 2015). One of them, the traffic ATPase VirB11, was shown to be the target for inhibition by unsaturated fatty acids (Ripoll-Rozada et al., 2016). Here, we will analyze the progress on the different strategies to inhibit the VirB11 ATPase and the rest of the T4SS machinery. The impact of these results on the fight against the spread of antibiotic resistance genes is discussed.

## Strategies for the Identification of Conjugation Inhibitors

Bacterial conjugation has been reported to be inhibited by a variety of compounds. Indeed, chemicals such as heterocyclic compounds, intercalators, acridine dyes, or quinolones were reported to inhibit conjugation (Hahn and Ciak, 1976; Michel-Briand and Laporte, 1985; Molnar et al., 1992; Mazel and Davies, 1999; Nash et al., 2012). However, posterior revisions showed that these molecules were unspecific, mainly affecting bacterial growth or DNA synthesis. Plants are a rich source of bioactive compounds, such as phenolics, which are able to modify bacterial resistances (Oyedemi et al., 2016). Therefore, a current approach consists of isolating molecules from different parts of medicinal plants to discover new inhibitors. By using this approach, two new drugs: rottlerin [5,7-dihydroxy-2,2-dimethyl-6-(2,4,6-trihydroxy-3-methyl-5-acetylbenzyl)-8-cinnamoyl-1,2-chromene] and the red compound (8-cinnamoyl-5,7-dihydroxy-2,2,6-trimethylchromene) were identified as potent antibacterial chemicals against Gram-positive bacteria. These compounds did not hamper Gram-negative bacteria growth but inhibited

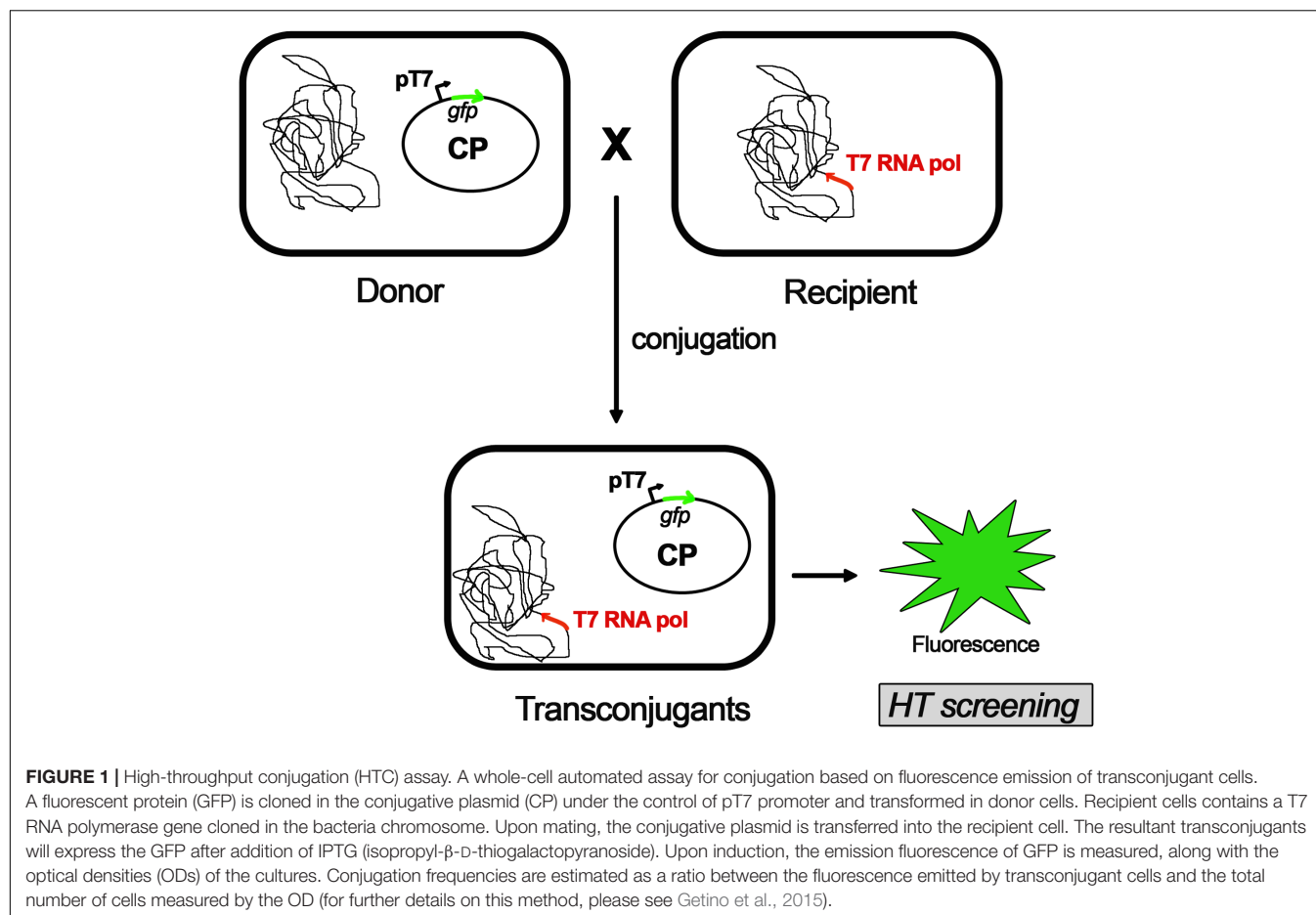
conjugal transfer of plasmids pKM101, TP114, pUB307, and R6K (Oyedemi et al., 2016). The planar structure of the compounds suggests that the target of these inhibitors might be the DNA replication system but further studies are required to elucidate the mode of inhibition of these agents.

Alternative attempts to inhibit bacterial conjugation have been based on bottom up strategies, targeting essential compounds of the secretion machinery. One study focused on targeting the conjugative relaxase protein, which is the protein that initiates conjugation upon nicking plasmid DNA at the origin of transfer. Due to its key role in plasmid conjugation, relaxases have been considered as potential targets for inhibitors. Some of these potential relaxase-specific inhibitors belong to the bisphosphonates family of compounds, such as etidronate (Didronel) and clodronate (Bonefos) (Lujan et al., 2007). These compounds were reported to be efficient in restraining conjugative DNA transfer. However, these results turned out to be misleading, as these putative inhibitors were found to work as unspecific chelating agents (Nash et al., 2012). An alternative method to inhibit specifically the conjugative relaxase consisted of the expression of specific single chain Fv antibodies (intrabodies) against the relaxase TrwC of conjugative plasmid R388 (Garcillán-Barcia et al., 2007). Expression of these intrabodies in the recipient cell prevented the accretion of the conjugative plasmid. However, the usefulness of intrabodies in practical clinical care is hampered by the need of a transgenic recipient population expressing them. Besides, each intrabody would be specific only against its cognate plasmid.

VirB8 is an essential assembly protein of bacterial T4SS that also acts as molecular target of small-molecule inhibitors (Smith et al., 2012). A high throughput assay based on the restoration of interactions between two split domains of the *Brucella* VirB8 protein allowed the identification of several compounds that inhibited protein-protein interactions (Paschos et al., 2011). One of the most efficient molecules, B8I-2, is a salicylidene acyl-hydrazide derivative, also known to inhibit T3SS (Keyser et al., 2008). Posterior analysis by X-ray crystallography and *in silico* docking of several of these compounds allowed the determination of VirB8 binding site (Smith et al., 2012). Recently, it has been reported that these small molecules also bind TraE, the VirB8 homolog of the conjugative plasmid pKM101, and some of them inhibit plasmid transfer (Casu et al., 2016). Although some of these molecules displayed a low K<sub>d</sub> value in *in vitro* binding experiments, no significant impact was observed on plasmid transfer frequencies, with a 10-fold reduction as the strongest effect. Moreover, none of these molecules had an effect on the conjugation of the unrelated plasmid RP4, diminishing the effectiveness of these compounds to overcome antibiotic resistance.

Other alternatives to develop specific inhibitors focused on the conjugative pilus. These appendages are targeted by a variety of bacteriophages that, upon binding to them, enter inside the bacteria cytoplasm. Some of these bacteriophages are specific to conjugative pili and, therefore, have the potential of discriminating between different bacterial species. For instance, filamentous bacteriophages, such as M13, display high affinity for F-type pilus. This interaction is mediated by the phage





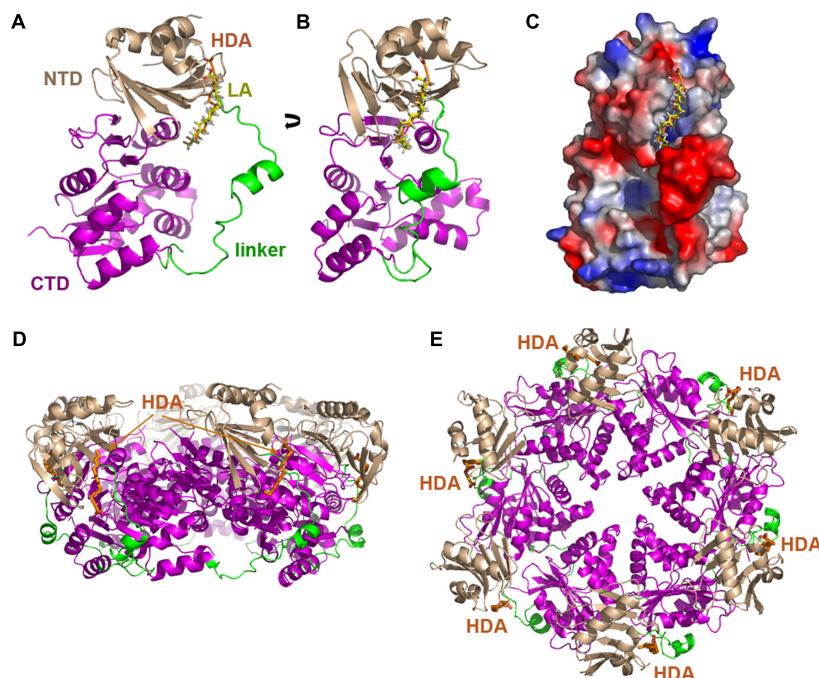
coat protein g3p. Addition of the soluble N-terminal domain of g3p to F-plasmid containing bacteria resulted in inhibition of conjugation (Lin et al., 2011). Considering that conjugative pili are needed for bacterial cell contact (Anthony et al., 1994), a step required for the spread of antibiotic resistance genes, the use of specific compounds that inhibit pilus formation could result in a powerful strategy to fight against this problem.

In that sense, peptidomimetic small molecules, such as C10 and KSK85, have been found to disrupt T4SS-dependent transport of pathogenic factors, as well as DNA transfer in conjugative *Escherichia coli*. KSK85 acts impeding biogenesis of the pilus appendage, whereas C10 disrupts T4SS activity without affecting pilus assembly (Shaffer et al., 2016). In this case, authors have used a phenotypic screen to identify these two inhibitors. Both compounds have been tested in plasmids pKM101 (IncN) and R1-16 (IncF) but conjugation efficiency is only decreased to 25% and a high concentration of inhibitor is required (150 μM). Therefore, although these compounds are promising scaffolds, new derivatives need to be found to inhibit more effectively the conjugative process.

A novel approach in the pursuit of specific COINs is the screening of potential compounds in whole-cell assays (Figure 1). This methodology should be designed in a way that

discriminates COINs from false positives affecting cell growth. Thus, the development of luminescence-based high-throughput conjugation (HTC) assays has been shown to be effective in the identification of potential hits. By using HTC assays, unsaturated fatty acids were found to inhibit conjugation of IncF and IncW plasmids without affecting cell growth (Fernández-Lopez et al., 2005). To this end, a library consisting of more than 12,000 natural compounds (NatChem library) was tested. The most effective COIN found in this screening was dehydrocrepynic acid (DHCA) (Fernández-Lopez et al., 2005). Considering that this compound was extracted from tropical plant seeds (Gussoni et al., 1994) the viability of using it without vast downstream process improvement is limited.

Nonetheless, by using DHCA structure as a chemical template, new synthetic compounds were developed as specific COINs. In particular, synthetic 2-hexadecynoic acid (2-HDA) and other 2-alkynoic fatty acids (2-AFAs) were found to be specific inhibitors of a wide range of conjugative plasmids in different bacteria, including the highly infective and prevalent IncF plasmids (Getino et al., 2015). Furthermore, due to the effect of plasmid burden on host fitness, 2-AFAs could in fact eliminate transmissible plasmids, such as IncF, from bacterial populations. However, other plasmid groups, such as IncN and IncP, were not affected (Fernández-Lopez et al., 2005; Getino et al., 2015).



**FIGURE 2 |** Blind docking of fatty acids in TrwD. Docking predictions between a molecular model of TrwD, the traffic ATPase of the conjugative plasmid R388, and linoleic acid and 2-HDA are shown (**A,B**) (Ripoll-Rozada et al., 2016). These unsaturated fatty acids bind to a pocket formed by the end of the N-terminal domain (NTD, *wheat*) and the linker region (*green*) that connects it to the C-terminal domain (CTD, *purple*). An electrostatic map showing these interactions is included in (**C**). Inhibitor binding sites in the hexameric TrwD protein are shown in (**D,E**), corresponding to side and top views, respectively. Docking was performed using the EADock dihedral spacing sampling engine of the Swiss-dock server (Grosdidier et al., 2011).

In addition to 2-AFAs compounds, another family of bioactive compounds was tested in the search for COINs. This collection, named AQUAc, contains more than 1,600 natural compounds. AQUAc was evaluated in order to find potential COINs (Getino et al., 2016). As a result, new COINs were found. Among them, tanzawaic acids A and B were identified as best hits. They specifically inhibited IncW and IncFII conjugative plasmids. The advantage of these compounds is their lower toxicity to animal cells, in comparison to other synthetic COINs. Unsaturated fatty acids (oleic and linoleic acids), 2-HDA, 2-alkynoic fatty acids and tanzawaic acids, all share similar chemical characteristics: a carboxylic group, a long unsaturated carbon chain and the presence of double or triple bonds. These compounds present a 100-fold reduction in plasmid transfer frequencies and, although higher inhibition rates must be achieved to maximize their effectiveness, they constitute key scaffold structures on which to develop more potent and specific COINs. In that respect, knowing the molecular target of these compounds is extremely important, since the use of structure-based drug design (SBDD) methods will allow the design of modified synthetic compounds with higher binding affinities.

## Unsaturated Fatty Acids As Specific Inhibitors of Conjugative Traffic ATPases

Bacterial conjugation is driven by a group of ATPases that empowers almost every step in the conjugative process:

DNA unwinding, DNA transport, pilus biogenesis and protein transport (Cabezón et al., 2015). Each of these steps is catalyzed by a specific ATPase, that in the conjugative plasmid R388 are named TrwC, TrwB, TrwK, and TrwD, respectively. TrwC is a protein that nicks the DNA, thus relaxing the conjugative plasmid in an ATP-independent manner (Llosa et al., 1995, 1996). In addition, TrwC displays a DNA helicase activity that results in nucleic acid unwinding (Grandoso et al., 1994; Llosa et al., 1996). TrwB is a DNA-dependent ATPase involved in DNA transfer to the secretion channel (Tato et al., 2005, 2007; Matilla et al., 2010). TrwK is a hexameric ATPase (Arechaga et al., 2008; Peña et al., 2011) that participates in the transport of the pilin molecules from the inner to the outer membranes during pilus biogenesis (Kerr and Christie, 2010). Finally, TrwD is a traffic ATPase that contributes to pilus biogenesis and to DNA translocation (Atmakuri et al., 2004), thus working as a molecular switch between pilus synthesis and substrate transport (Ripoll-Rozada et al., 2013).

Each of these ATPases has been purified to homogeneity and their enzymatic activities have been characterized (Grandoso et al., 1994; Tato et al., 2005; Arechaga et al., 2008; Ripoll-Rozada et al., 2012). The kinetic parameters of each ATPase were analyzed in the presence of the unsaturated fatty acids shown to be efficient COINs (Ripoll-Rozada et al., 2016). Interestingly, only the traffic ATPase TrwD was inhibited by unsaturated fatty acids, such a linoleic acid, and 2-AFAs like 2-HDA. The kinetic parameters for TrwD ATPase inhibition by these fatty acids were determined,

indicating that in all cases it was a non-competitive inhibition (Ripoll-Rozada et al., 2016). In contrast, saturated fatty acids, such as palmitic acid, showed no inhibitory effect in conjugation experiments and in ATPase assays.

TrwD belongs to the secretion ATPase superfamily, which also includes members of Type II secretion, Type IV pilus and flagellar biogenesis machineries (Planet et al., 2001). All members of this superfamily are hexameric ATPases, in which each monomer is formed by two domains at the N- and C-termini (NTD and CTD, respectively), connected by a flexible linker of a variable length (Planet et al., 2001; Peña and Arechaga, 2013). ATPase catalysis is driven by swapping this linker over the NTD and CTD (Savvides et al., 2003; Hare et al., 2006). Blind docking predictions (Grosdidier et al., 2011) suggested a putative binding site for uFAs and 2-aFAs located at the end of the NTD and beginning of the linker region that connects it to the CTD where the nucleotide binding site is located (Figure 2) (Ripoll-Rozada et al., 2016). These predictions were compatible with a model in which the mode of action of the inhibitors consisted in preventing the swapping movements between the N- and C-terminal domains along the linker region that are required in the catalytic cycle of the protein.

In summary, the discovery of traffic ATPases as potential targets of bacterial COINs opens a promising avenue for the development of new and more potent compounds capable to impair the dissemination of antibiotic resistance genes. However, despite these promising data, the efficacy of these inhibitors to prevent the general spread of antibiotic resistance in naturally occurring environments (Martinez, 2008; von Wintersdorff et al., 2016), hospitals, wastewater systems (Hocquet et al., 2016), agriculture (Capita and Alonso-Calleja, 2013; Economou and Gousia, 2015) and farming settings (Wegener, 2003) needs to be tested. Even today, the main reservoirs where the antibiotic resistance genes arise from and how these genes are rapidly acquired by human pathogens are a matter of debate. COINs could be effective in the discovery of these natural reservoirs. Experiments in controlled microcosms (e.g., freshwater microcosms) and/or experimental animals (e.g., mice gut) should be most instructive. Not only COIN-related experiments will help to identify reservoirs, they will also serve to quantify the dynamics of plasmids in those experiments and the rate at which they can be mobilized to recipient strains that

are potential human pathogens. These are some of the exciting goals of ongoing COINs research.

## OUTLOOK

The battle against antibiotic resistance is a challenging problem which is likely to become a progressively increasing burden to our health systems. Several approaches are currently envisaged to fight back against antibiotic resistant bugs. Among them, a promising alternative consists of preventing the propagation of antibiotic resistance genes. Bacterial conjugation is the main mechanism for the wide spread dissemination of these genes. Hence, the search of compounds able to specifically inhibit bacterial conjugation is a preeminent undertaking in the global war against antibiotic resistant bugs. Here, we have reviewed several compounds that are competent in this pursuit. Among these, unsaturated fatty acids and derivatives have been proved to be the most efficient specific COINs. Moreover, the identification of the traffic ATPase TrwD as the molecular target of these uFAs enables the future development of more efficient inhibitors directed against this essential protein of T4SS.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This work was supported by the Spanish Ministerio de Economía y Competitividad (MINECO) grants BFU2016-78521-R (to EC and IA) and BFU2014-55534-C2-1-P (to fdlc).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02329/full#supplementary-material>

## REFERENCES

- Anthony, K. G., Sherburne, C., Sherburne, R., and Frost, L. S. (1994). The role of the pilus in recipient cell recognition during bacterial conjugation mediated by F-like plasmids. *Mol. Microbiol.* 13, 939–953. doi: 10.1111/j.1365-2958.1994.tb00486.x
- Arechaga, I., Peña, A., Zunzunegui, S., del Carmen Fernandez-Alonso, M., Rivas, G., and de la Cruz, F. (2008). ATPase activity and oligomeric state of TrwK, the VirB4 homologue of the plasmid R388 type IV secretion system. *J. Bacteriol.* 190, 5472–5479. doi: 10.1128/JB.00321-08
- Atmakuri, K., Cascales, E., and Christie, P. J. (2004). ). Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Mol. Microbiol.* 54, 1199–1211. doi: 10.1111/j.1365-2958.2004.04345.x
- Baquero, F., Coque, T. M., and de la Cruz, F. (2011). Ecology and evolution as targets: the need for novel eco-evo drugs and strategies to fight antibiotic resistance. *Antimicrob. Agents Chemother.* 55, 3649–3660. doi: 10.1128/AAC.00013-11
- Baym, M., Stone, L. K., and Kishony, R. (2016). Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* 351:aad3292. doi: 10.1126/science.aad3292
- Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., et al. (2009). Bad bugs, no drugs: no ESKAPE! An update from the infectious diseases society of America. *Clin. Infect. Dis.* 48, 1–12. doi: 10.1086/595011
- Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* 39, 81–95. doi: 10.1111/1574-6976.12085
- Capita, R., and Alonso-Calleja, C. (2013). Antibiotic-resistant bacteria: a challenge for the food industry. *Crit. Rev. Food Sci. Nutr.* 53, 11–48. doi: 10.1080/10408398.2010.519837

- Casu, B., Smart, J., Hancock, M. A., Smith, M., Sygusch, J., and Baron, C. (2016). Structural analysis and inhibition of TraE from the pKM101 type IV secretion system. *J. Biol. Chem.* 291, 23817–23829. doi: 10.1074/jbc.M116.753327
- Chang, Q., Wang, W., Regev-Yochay, G., Lipsitch, M., and Hanage, W. P. (2015). Antibiotics in agriculture and the risk to human health: how worried should we be? *Evol. Appl.* 8, 240–247. doi: 10.1111/eva.12185
- Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S., and Cascales, E. (2005). Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu. Rev. Microbiol.* 59, 451–485. doi: 10.1146/annurev.micro.58.030603.123630
- Christie, P. J., Whitaker, N., and Gonzalez-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta* 1843, 1578–1591. doi: 10.1016/j.bbamcr.2013.12.019
- Conter, A., Sturny, R., Gutierrez, C., and Cam, K. (2002). The RcsCB His-Asp phosphorelay system is essential to overcome chlorpromazine-induced stress in *Escherichia coli*. *J. Bacteriol.* 184, 2850–2853. doi: 10.1128/JB.184.10.2850-2853.2002
- de la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.* 34, 18–40. doi: 10.1111/j.1574-6976.2009.00195.x
- Economou, V., and Gousia, P. (2015). Agriculture and food animals as a source of antimicrobial-resistant bacteria. *Infect. Drug Resist.* 8, 49–61. doi: 10.2147/IDR.S55778
- Fernandez-Lopez, R., Garcillan-Barcia, M. P., Revilla, C., Lazaro, M., Vielva, L., and de la Cruz, F. (2006). Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol. Rev.* 30, 942–966. doi: 10.1111/j.1574-6976.2006.00042.x
- Fernández-Lopez, R., Machon, C., Longshaw, C. M., Martin, S., Molin, S., Zechner, E. L., et al. (2005). Unsaturated fatty acids are inhibitors of bacterial conjugation. *Microbiology* 151(Pt 11), 3517–3526. doi: 10.1099/mic.0.28216-0
- Garcillán-Barcia, M. P., Francia, M. V., and de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* 33, 657–687. doi: 10.1111/j.1574-6976.2009.00168.x
- Garcillan-Barcia, M. P., Jurado, P., Gonzalez-Perez, B., Moncalian, G., Fernandez, L. A., and de la Cruz, F. (2007). Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. *Mol. Microbiol.* 63, 404–416. doi: 10.1111/j.1365-2958.2006.05523.x
- Getino, M., Sanabria-Ríos, D. J., Fernández-López, R., Campos-Gómez, J., Sánchez-López, J. M., Fernández, A., et al. (2015). Synthetic fatty acids prevent plasmid-mediated horizontal gene transfer. *mbio* 6:e01032-15. doi: 10.1128/mBio.01032-15
- Getino, M., Fernandez-Lopez, R., Palencia-Gandara, C., Campos-Gomez, J., Sanchez-Lopez, J. M., Martinez, M., et al. (2016). Tanzawaic acids, a chemically novel set of bacterial conjugation inhibitors. *PLOS ONE* 11:e0148098. doi: 10.1371/journal.pone.0148098
- Giske, C. G., Monnet, D. L., Cars, O., and Carmeli, Y. (2008). Clinical and economic impact of common multidrug-resistant gram-negative bacilli. *Antimicrob. Agents Chemother.* 52, 813–821. doi: 10.1128/AAC.01169-07
- Gould, I. M., and Bal, A. M. (2013). New antibiotic agents in the pipeline and how they can help overcome microbial resistance. *Virulence* 4, 185–191. doi: 10.4161/viru.22507
- Grandoso, G., Llosa, M., Zabala, J. C., and de la Cruz, F. (1994). Purification and biochemical characterization of TrwC, the helicase involved in plasmid R388 conjugal DNA transfer. *Eur. J. Biochem.* 226, 403–412. doi: 10.1111/j.1432-1033.1994.tb20065.x
- Grosdidier, A., Zoete, V., and Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* 39, W270–W277. doi: 10.1093/nar/gkr366
- Gussoni, M., Greco, F., Pegna, M., Bianchi, G., and Zetta, L. (1994). Solid state and microscopy NMR study of the chemical constituents of *Azelia cuanzensis* seeds. *Magn. Reson. Imaging* 12, 477–486. doi: 10.1016/0730-725X(94)92542-9
- Hahn, F. E., and Ciak, J. (1976). Elimination of resistance determinants from R-factor R1 by intercalative compounds. *Antimicrob. Agents Chemother.* 9, 77–80. doi: 10.1128/AAC.9.1.77
- Hare, S., Bayliss, R., Baron, C., and Waksman, G. (2006). A large domain swap in the VirB11 ATPase of *Brucella suis* leaves the hexameric assembly intact. *J. Mol. Biol.* 360, 56–66. doi: 10.1016/j.jmb.2006.04.060
- Hede, K. (2014). Antibiotic resistance: An infectious arms race. *Nature* 509, S2–S3. doi: 10.1038/509S2a
- Hocquet, D., Muller, A., and Bertrand, X. (2016). What happens in hospitals does not stay in hospitals: antibiotic-resistant bacteria in hospital waste water systems. *J. Hosp. Inf.* 93, 395–402. doi: 10.1016/j.jhin.2016.01.010
- Hooper, D. C., Wolfson, J. S., Tung, C., Souza, K. S., and Swartz, M. N. (1989). Effects of inhibition of the B subunit of DNA gyrase on conjugation in *Escherichia coli*. *J. Bacteriol.* 171, 2235–2237. doi: 10.1128/jb.171.4.2235-2237.1989
- Kerr, J. E., and Christie, P. J. (2010). Evidence for VirB4-mediated dislocation of membrane-integrated VirB2 pilin during biogenesis of the *Agrobacterium* VirB/VirD4 type IV secretion system. *J. Bacteriol.* 192, 4923–4934. doi: 10.1128/JB.00557-10
- Keyser, P., Elofsson, M., Rosell, S., and Wolf-Watz, H. (2008). Virulence blockers as alternatives to antibiotics: type III secretion inhibitors against gram-negative bacteria. *J. Intern. Med.* 264, 17–29. doi: 10.1111/j.1365-2796.2008.01941.x
- Lin, A., Jimenez, J., Derr, J., Vera, P., Manapat, M. L., Esvelt, K. M., et al. (2011). Inhibition of bacterial conjugation by phage M13 and its protein g3p: quantitative analysis and model. *PLOS ONE* 6:e19991. doi: 10.1371/journal.pone.0019991
- Llosa, M., Gomis-Ruth, F. X., Coll, M., and de la Cruz Fd, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* 45, 1–8. doi: 10.1046/j.1365-2958.2002.03014.x
- Llosa, M., Grandoso, G., and de la Cruz, F. (1995). Nicking activity of TrwC directed against the origin of transfer of the IncW plasmid R388. *J. Mol. Biol.* 246, 54–62. doi: 10.1006/jmbi.1994.0065
- Llosa, M., Grandoso, G., Hernando, M. A., and de la Cruz, F. (1996). Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. *J. Mol. Biol.* 264, 56–67. doi: 10.1006/jmbi.1996.0623
- Lujan, S. A., Guogas, L. M., Ragonese, H., Matson, S. W., and Redinbo, M. R. (2007). Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12282–12287. doi: 10.1073/pnas.0702760104
- Martinez, J. L. (2008). Antibiotics and antibiotic resistance genes in natural environments. *Science* 321, 365–367. doi: 10.1126/science.1159483
- Matilla, I., Alfonso, C., Rivas, G., Bolt, E. L., de la Cruz, F., and Cabezon, E. (2010). The conjugative DNA translocase TrwB is a structure-specific DNA-binding protein. *J. Biol. Chem.* 285, 17537–17544. doi: 10.1074/jbc.M109.084137
- Mazel, D., and Davies, J. (1999). Antibiotic resistance in microbes. *Cell Mol. Life Sci.* 56, 742–754. doi: 10.1007/s00180050021
- Michel-Briand, Y., and Laporte, J. M. (1985). Inhibition of conjugal transfer of R plasmids by nitrofurans. *J. Gen. Microbiol.* 131, 2281–2284. doi: 10.1099/00221287-131-9-2281
- Molnar, J., Fischer, J., and Nakamura, M. J. (1992). Mechanism of chlorpromazine binding by gram-positive and gram-negative bacteria. *Antonie Van Leeuwenhoek* 62, 309–314. doi: 10.1007/BF00572599
- Moncalian, G., and de la Cruz, F. (2004). DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. *Biochim. Biophys. Acta* 1701, 15–23. doi: 10.1016/j.bbapap.2004.05.009
- Nash, R. P., McNamara, D. E., Ballentine, W. K. III, Matson, S. W., and Redinbo, M. R. (2012). Investigating the impact of bisphosphonates and structurally related compounds on bacteria containing conjugative plasmids. *Biochem. Biophys. Res. Commun.* 424, 697–703. doi: 10.1016/j.bbrc.2012.07.012
- Oyedemi, B. O., Shinde, V., Shinde, K., Kakalou, D., Stapleton, P. D., and Gibbons, S. (2016). Novel R-plasmid conjugal transfer inhibitory and antibacterial activities of phenolic compounds from *Mallotus philippensis* (Lam.) Mull. Arg. *J. Glob. Antimicrob. Resist.* 5, 15–21. doi: 10.1016/j.jgar.2016.01.011
- Paschos, A., den Hartigh, A., Smith, M. A., Atluri, V. L., Sivanesan, D., Tsolis, R. M., et al. (2011). An *in vivo* high-throughput screening approach targeting the type IV secretion system component VirB8 identified inhibitors of *Brucella abortus* 2308 proliferation. *Infect. Immun.* 79, 1033–1043. doi: 10.1128/IAI.00993-10
- Peña, A., and Arechaga, I. (2013). Molecular motors in bacterial secretion. *J. Mol. Microbiol. Biotechnol.* 23, 357–369. doi: 10.1159/000351360
- Peña, A., Ripoll-Rozada, J., Zunzunegui, S., Cabezon, E., de la Cruz, F., and Arechaga, I. (2011). Autoinhibitory regulation of TrwK, an essential VirB4



- ATPase in type IV secretion systems. *J. Biol. Chem.* 286, 17376–17382. doi: 10.1074/jbc.M110.208942
- Planet, P. J., Kachlany, S. C., DeSalle, R., and Figurski, D. H. (2001). Phylogeny of genes for secretion NTPases: identification of the widespread *tadA* subfamily and development of a diagnostic key for gene classification. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2503–2508. doi: 10.1073/pnas.051436598
- Rice, L. B. (2009). The clinical consequences of antimicrobial resistance. *Curr. Opin. Microbiol.* 12, 476–481. doi: 10.1016/j.mib.2009.08.001
- Ripoll-Rozada, J., García-Cazorla, Y., Getino, M., Machon, C., Sanabria-Rios, D., de la Cruz, F., et al. (2016). Type IV traffic ATPase TrwD as molecular target to inhibit bacterial conjugation. *Mol. Microbiol.* 100, 912–921. doi: 10.1111/mmi.13359
- Ripoll-Rozada, J., Pena, A., Rivas, S., Moro, F., de la Cruz, F., Cabezón, E., et al. (2012). Regulation of the type IV secretion ATPase TrwD by magnesium: implications for catalytic mechanism of the secretion ATPase superfamily. *J. Biol. Chem.* 287, 17408–17414. doi: 10.1074/jbc.M112.357905
- Ripoll-Rozada, J., Zunzunegui, S., de la Cruz, F., Arechaga, I., and Cabezón, E. (2013). Functional interactions of VirB11 traffic ATPases with VirB4 and VirD4 molecular motors in type IV secretion systems. *J. Bacteriol.* 195, 4195–4201. doi: 10.1128/JB.00437-13
- Savvides, S. N., Yeo, H. J., Beck, M. R., Blaesing, F., Lurz, R., Lanka, E., et al. (2003). VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion. *EMBO J.* 22, 1969–1980. doi: 10.1093/emboj/cdg223
- Shaffer, C. L., Good, J. A. D., Kumar, S., Krishnan, K. S., Gaddy, J. A., Loh, J. T., et al. (2016). Peptidomimetic small molecules disrupt type IV secretion activity in diverse bacterial pathogens. *mBio* 7:e221-16. doi: 10.1128/mBio.00221-16
- Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10
- Smith, M. A., Coincon, M., Paschos, A., Jolicoeur, B., Lavallee, P., Sygusch, J., et al. (2012). Identification of the binding site of *Brucella* VirB8 interaction inhibitors. *Chem. Biol.* 19, 1041–1048. doi: 10.1016/j.chembiol.2012.07.007
- Smith, P. A., and Romesberg, F. E. (2007). Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nat. Chem. Biol.* 3, 549–556. doi: 10.1038/nchembio.2007.27
- Tato, I., Matilla, I., Arechaga, I., Zunzunegui, S., de la Cruz, F., and Cabezón, E. (2007). The ATPase activity of the DNA transporter TrwB is modulated by protein TrwA: implications for a common assembly mechanism of DNA translocating motors. *J. Biol. Chem.* 282, 25569–25576. doi: 10.1074/jbc.M703464200
- Tato, I., Zunzunegui, S., de la Cruz, F., and Cabezón, E. (2005). TrwB, the coupling protein involved in DNA transport during bacterial conjugation, is a DNA-dependent ATPase. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8156–8161. doi: 10.1073/pnas.0503402102
- von Wintersdorff, C., Penders, J., van Niekerk, J. M., Mills, N. D., Majumder, S., van Alphen, L. B., et al. (2016). Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front. Microbiol.* 7:173. doi: 10.3389/fmicb.2016.00173
- Waters, V. L. (1999). Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance. *Front. Biosci.* 4, D433–D456.
- Wegener, H. C. (2003). Antibiotics in animal feed and their role in resistance development. *Curr. Opin. Microbiol.* 6, 439–445. doi: 10.1016/j.mib.2003.09.009
- WHO (2014). *WHO: Antimicrobial Resistance. Global Report on Surveillance*. Geneva: WHO.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Cabezón, de la Cruz and Arechaga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Novel Antimicrobial Coating Represses Biofilm and Virulence-Related Genes in Methicillin-Resistant *Staphylococcus aureus*

Ankita Vaishampayan<sup>1</sup>, Anne de Jong<sup>2</sup>, Darren J. Wight<sup>3</sup>, Jan Kok<sup>2</sup> and Elisabeth Grohmann<sup>1,4\*</sup>

<sup>1</sup> Life Sciences and Technology, Beuth University of Applied Sciences Berlin, Berlin, Germany, <sup>2</sup> Department of Molecular Genetics, University of Groningen, Groningen, Netherlands, <sup>3</sup> Institute of Virology, Free University of Berlin, Berlin, Germany, <sup>4</sup> Division of Infectious Diseases, University Medical Center Freiburg, Freiburg, Germany

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biomédicas  
(CSIC), Spain

### Reviewed by:

Mirian Moscoso,  
Instituto de Investigación Biomédica  
de A Coruña (INIBIC), Spain  
María Victoria Francia,  
Marqués de Valdecilla University  
Hospital, Spain

### \*Correspondence:

Elisabeth Grohmann  
elisabeth.grohmann@  
beuth-hochschule.de

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 29 September 2017

**Accepted:** 30 January 2018

**Published:** 15 February 2018

### Citation:

Vaishampayan A, de Jong A,  
Wight DJ, Kok J and Grohmann E  
(2018) A Novel Antimicrobial Coating  
Represses Biofilm  
and Virulence-Related Genes  
in Methicillin-Resistant  
*Staphylococcus aureus*.  
Front. Microbiol. 9:221.  
doi: 10.3389/fmicb.2018.00221

Methicillin-resistant *Staphylococcus aureus* (MRSA) has become an important cause of hospital-acquired infections worldwide. It is one of the most threatening pathogens due to its multi-drug resistance and strong biofilm-forming capacity. Thus, there is an urgent need for novel alternative strategies to combat bacterial infections. Recently, we demonstrated that a novel antimicrobial surface coating, AGXX®, consisting of micro-galvanic elements of the two noble metals, silver and ruthenium, surface-conditioned with ascorbic acid, efficiently inhibits MRSA growth. In this study, we demonstrated that the antimicrobial coating caused a significant reduction in biofilm formation (46%) of the clinical MRSA isolate, *S. aureus* 04-02981. To understand the molecular mechanism of the antimicrobial coating, we exposed *S. aureus* 04-02981 for different time-periods to the coating and investigated its molecular response via next-generation RNA-sequencing. A conventional antimicrobial silver coating served as a control. RNA-sequencing demonstrated down-regulation of many biofilm-associated genes and of genes related to virulence of *S. aureus*. The antimicrobial substance also down-regulated the two-component quorum-sensing system *agr* suggesting that it might interfere with quorum-sensing while diminishing biofilm formation in *S. aureus* 04-02981.

**Keywords:** antimicrobial surface, MRSA, virulence, biofilm, quorum-sensing, RNA sequencing

## INTRODUCTION

*Staphylococcus aureus* is an opportunistic pathogen commonly found in the human respiratory tract, nasal areas and skin. It colonizes the anterior nares of approximately 20–25% of the healthy adult population, while 60% are intermittently colonized (Kluytmans et al., 1997; Ellis et al., 2014). Methicillin-resistant *Staphylococcus aureus* (MRSA) is a crucial human pathogen causing infections ranging from skin and soft tissue infections to fatal sepsis (Marathe et al., 2015). It is one of the leading pathogens that cause nosocomial infections (Paniagua-Contreras et al., 2012; Lister and Horswill, 2014); it is resistant to methicillin and many other antibiotics (Marathe et al., 2015), and it is also known to produce thick biofilm (Paniagua-Contreras et al., 2012; Qin et al., 2014). MRSA was shown to cause catheter-associated and other medical devices-related

infections (Arciola et al., 2001; Paniagua-Contreras et al., 2012). Eighty percent of prosthetic infections are caused by Staphylococci (Kirmusaoglu, 2016). Its firm attachment to medical devices and host tissues, and its ability to form robust biofilms makes it a cause of chronic infections (Yarwood et al., 2004). *S. aureus* biofilms cause numerous infections in which the accessory gene regulator (*agr*) quorum-sensing system (QS) plays an important role (Yarwood et al., 2004). Around 90% of the infections caused by the bacterium are skin and soft tissue infections, and the *agr*QS system is associated with these infections (Sully et al., 2014).

Multiple drug resistance combined with a thick biofilm makes the treatment and eradication of *S. aureus* infections even more difficult. This entails the urge of development of novel antimicrobials, which could also be potential biofilm inhibitors. Virulence factors of *S. aureus* serve as targets for the newly developed class of biological anti-staphylococcal agents. These targets include, surface bound adhesins, immunoglobulin-binding proteins, surface-associated and secreted proteases, a family of immune-stimulatory exotoxins called 'superantigens' (SAGs), and potent leukocidal toxins (Sause et al., 2015).

Metals like copper and silver have been used as antimicrobials since a long time. The use of copper in human civilization is known since the 5th and 6th millennia B.C. Silver was officially approved for use as an antimicrobial agent in the 20th century (Chopra, 2007; Grass et al., 2011; Schäberle and Hack, 2014; Guridi et al., 2015). Copper and copper alloys have also been used as antimicrobials (Warnes and Keevil, 2013). These metals are known to kill bacteria and fungi by a phenomenon called contact killing (Grass et al., 2011) and can be used to coat medical devices as they inhibit biofilm formation of pathogens (Baker et al., 2010). In the 17th century, silver was described as an essential multipurpose medicinal product and the first scientific documentation of its medical use dates from 1901 (Maillard and Hartemann, 2013). However, in 1975, several patients died from a silver resistant *Salmonella* Typhimurium isolate in the Massachusetts General Hospital; this was the first report of silver resistant bacteria (Gupta et al., 1999). Excessive use of silver is questioned due to its toxicity to the environment as well as to the human body (Landsdown, 2010). Silver resistance, like antibiotic resistance in bacteria, prompts us to develop new strategies to control bacterial infections. One such novel, broad-spectrum antimicrobial agent is AGXX®.

AGXX® (Largentec GmbH, Berlin, Germany) is a combination of two transition metals, silver and ruthenium which can be galvanically electroplated on various carriers like V2A steel, silver sheets, Polydimethylsiloxane (PDMS), fleece, etc. The coating is conditioned by ascorbic acid and is active against many Gram-positive and Gram-negative bacteria (Guridi et al., 2015). It is not only an efficient antibacterial but also kills yeasts, viruses, and fungi (Landau et al., 2017a,b). The coating was used successfully for the decontamination of industrial cooling and process water (Landau, 2013). As it is only slightly cytotoxic (Bouchard, 2011), it can be incorporated into various medical applications. Although, the exact mode of action of the antimicrobial activity of the coating is not fully understood, it is known that the generation of reactive oxygen species (ROS)

plays an important role in making it a potent antimicrobial. The formation of hydrogen peroxide and hydroxyl radicals has been detected by spectroscopic methods (Clauss-Lendzian et al., 2017). Putative formation of other ROS is under investigation. ROS can damage cellular components, including, DNA, lipids and proteins. Superoxide dismutase and catalase are involved in detoxification of ROS (Paraje, 2011).

In this study, we performed total RNA-sequencing of *S. aureus* 04-02981 (MRSA) to investigate differential gene expression after different times of exposure of the pathogen to the antimicrobials AGXX® or Ag. Our data demonstrate that AGXX® likely reduces biofilm formation and virulence in *S. aureus* 04-02981 by interfering with the QS, by down-regulating the expression of toxins like leukocidins (*lukE*) and gamma-hemolysins (*hlgA*), and of genes associated with surface adhesins and capsular polysaccharide.

## MATERIALS AND METHODS

### Preparation of Antimicrobial Metal Sheets

Silver sheets of 0.125 mm thickness were used as a base material to prepare the antimicrobial metal sheets. Both sides of the silver sheets were etched by immersing them in half-concentrated nitric acid, for 60 s. The silver sheets were cleaned with de-ionized water and galvanically plated with a 0.16 µm ruthenium coating on both sides for 40 s. Then, the sheets were cleaned with de-ionized water, conditioned with ascorbic acid, rinsed with de-ionized water and dried with a paper towel. Prior to use, AGXX®, and Ag sheets, used as reference material, were autoclaved at 121°C for 20 min.

### Bacterial Strain and Culture Conditions

*Staphylococcus aureus* 04-02981 (Nuebel et al., 2010) was grown at 37°C in Tryptic Soy Broth [TSB] (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) with constant agitation at 150 rpm or on Tryptic Soy Agar [TSA] (Carl Roth GmbH & Co. KG, Karlsruhe, Germany). Growth inhibition tests on agar surface were performed according to CLSI guidelines for disk diffusion test (Naas et al., 2006). For this assay, 0.25 cm<sup>2</sup> sheets of Ag and AGXX® were used.

For generation of growth curves, bacteria were pre-cultured overnight, diluted in TSB to an optical density at 600 nm (OD<sub>600</sub>) of 0.05 and incubated for further 8 h either in presence of AGXX® or in the presence of silver (Ag), 24 cm<sup>2</sup> each in 30 mL medium to obtain a sheet surface to medium volume ratio (A: V) of 0.8. Cultures grown in the absence of a metal sheet served as controls. OD<sub>600</sub> of the cultures was measured using the Genesys 10S UV-Vis spectrophotometer (Thermo Scientific, China). Colony forming units (CFU) per mL were determined hourly from 0 to 8 h post inoculation. Growth experiments were performed in triplicate with independent biological replicates.

### Biofilm Screening Assay

To study the effect of Ag, and AGXX® on biofilm formation of *S. aureus* 04-02981, the Crystal Violet Assay was performed

without any metal sheet, in presence of Ag (24 cm<sup>2</sup> uncoated silver sheet) and in presence of AGXX® (24 cm<sup>2</sup> silver sheet coated with ruthenium for 40 s). The sheet surface: medium volume ratio (A: V) was 0.8 (24 cm<sup>2</sup> metal sheet: 30 mL medium). The overnight culture of *S. aureus* 04-02981 was diluted to an initial OD<sub>600</sub> of 0.05. The culture was incubated at 37°C and 150 rpm for 4 h (mid-exponential phase, OD<sub>600</sub> ~1.5). Then, it was transferred to the transparent 96-well plate (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) containing Ag, or AGXX®. The plate was incubated at 37°C for 24 h, then the antimicrobial metal sheets were carefully removed and OD<sub>600</sub> of the cultures was measured. In addition, at this stage, the CFU per mL of the planktonic cultures and the biofilms in presence as well as in absence of the metal sheets were determined. Means of five values each and two biological replicates are given. The biofilm assay was performed according to Schiwon et al. (2013). *Enterococcus faecalis* 12030, a strong biofilm former was used as a positive control (Huebner et al., 1999), and Tryptic Soy Broth (TSB) as a negative control (Schiwon et al., 2013). Biofilm formation was measured in EnSpire Multimode Plate Reader 2300-0000 (PerkinElmer, Turku, Finland) at 570 nm. Normalized biofilm formation was calculated by dividing the biofilm measure at OD<sub>570</sub> by the bacterial growth at OD<sub>600</sub>. Following criteria were used for the interpretation of the results, OD<sub>c</sub> = negative control; OD ≤ OD<sub>c</sub> = non-adherent, OD<sub>c</sub> ≤ OD ≤ (2 × OD<sub>c</sub>) = weakly adherent, (2 × OD<sub>c</sub>) < OD ≤ (4 × OD<sub>c</sub>) = moderately adherent, (4 × OD<sub>c</sub>) < OD = strongly adherent, as described in Nyenje et al. (2013). Biofilm inhibitory rates of AGXX® and Ag were calculated using the following equation, as described by Qin et al. (2014).

$$\text{Inhibitory rate (\%)} = \frac{\text{OD}_{570}(\text{Control}) - \text{OD}_{570}(\text{Sample}) \times 100}{\text{OD}_{570}(\text{Control})}$$

Student's *t*-test was used to check if biofilm inhibition was statistically significant, using SigmaPlot version 11.0 (Systat software, Inc., San Jose, CA, United States<sup>1</sup>) (Wass, 2009).

## Spinning Disk Confocal Microscopy

*Staphylococcus aureus* 04-02981 was grown in TSB overnight at 37°C, 150 rpm, then it was diluted to an OD<sub>600</sub> of 0.05 and further incubated at 37°C for 4 h (mid-exponential phase, OD<sub>600</sub> ~1.5). Then, the culture was transferred to a μ-Dish (μ-Dish 35 mm, low, from ibidi GmbH, Martinsried, Germany) containing Ag, or AGXX® (sheet surface: medium volume ratio = 0.8) and incubated at 37°C for 24 h. The culture was removed from the μ-Dish, and the biofilm on the μ-Dish was washed three times with phosphate buffered saline (PBS). The biofilm was stained for 10 min in the dark with Hoechst 33342 (5 μg/mL) and propidium iodide (1 μg/mL) (Thermo Fisher, Eugene, OR, United States). The staining solution was then replaced with 50% glycerol to prevent movement of bacteria during imaging. Imaging was performed with a Nikon TiE-based Visitron spinning disk confocal microscope using a 100× NA1.45 objective. Fluorescent dyes were excited using 405 nm (Hoechst 33342) and 561 nm

(propidium iodide) laser lines and fluorescent emission captured through appropriate filters onto an iXon888 EMCCD detector (Andor, Belfast, United Kingdom). Images were subsequently analyzed using Fiji (ImageJ) version 3.2.0.2.

## Metal Stress and RNA Extraction

Overnight cultures of *S. aureus* 04-02981 were diluted as described above and grown until mid-exponential growth phase (4 h post dilution, OD<sub>600</sub> ~1.5). The cultures were then subjected to metal stress by exposure to AGXX® or Ag sheets (sheet-surface to medium-volume ratio of 0.8) followed by further incubation for 6, 12, 24, 80, and 120 min at 37°C with constant agitation at 150 rpm. As a control, no metal sheet was added to the culture. Cells from 30 mL culture were harvested by centrifugation for 1 min at 10,000 rpm and 4°C in a Heraeus Multifuge X3R Centrifuge (Thermo Electron LED GmbH, Osterode am Harz, Germany). Cell pellets were immediately frozen in liquid nitrogen and stored at −80°C or directly used for RNA extraction using the ZR Fungal/Bacterial RNA MiniPrep™ Kit (ZymoResearch, Freiburg, Germany) following the manufacturer's instructions. To recover total RNA including small RNAs, 1.5 volumes of absolute ethanol were added in step 5. Finally, total RNA was eluted with 50 μl DNase- and RNase-free water and stored at −80°C. RNA quantity and quality were assessed with a NanoDrop 2000c UV-Vis Spectrophotometer (Thermo Scientific, Osterode am Harz, Germany) as well as on bleach agarose gels. Residual contaminating DNA was eliminated with TURBO DNA-free™ Kit Ambion (Life Technologies, Darmstadt, Germany).

## RNA Sequencing

Total RNA sequencing was done by PrimBio Research Institute, Exton, PA, United States. The protocol was performed in five steps; rRNA removal was done using the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina, Cat# MRZMB126), followed by library preparation, and templating, enrichment and sequencing.

## RNA-Sequencing Data Analysis

Raw sequencing reads were aligned to the reference genome of *S. aureus* 04-02981, using Bowtie2 (Langmead and Salzberg, 2012) version 2.2.3 with optimal settings for the IonProton™ Sequence. Post-processing of the SAM files into sorted BAM files was carried out with SAMtools (Li et al., 2009, version 1.2-207). The samples AGXX®, and Ag were normalized (AGXX®-Control, Ag-Control) against the control of the respective time-points. Length normalized confidence interval RPKM (=Reads per Kilobase of transcript per Million mapped reads) values were obtained with Cufflinks (Trapnell et al., 2010). Finally, statistical analysis was carried out using the T-REX RNA-Seq analysis pipeline (de Jong et al., 2015). A gene was considered significantly differentially expressed when the fold change was ≥|2.0| and the false discovery rate (FDR) adjusted *p*-value ≤ 0.05. The data presented in this paper have been deposited at NCBI, and are accessible through GSE103064<sup>2</sup>.

<sup>1</sup><http://www.systatsoftware.com>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103064>



## Reverse Transcription Quantitative PCR (RT-qPCR)

To verify the results obtained from RNA-sequencing, RT-qPCR was performed on five genes detected as highly differentially expressed via RNA-seq. To this end, RNA extracted from *S. aureus* 04-02981 cultures exposed to Ag or AGXX® for 24, and 80 min, was used. First strand cDNA was synthesized with RevertAid™ First Strand cDNA Synthesis kit (Thermo Fisher Scientific Inc., Walham, Germany) as per the manufacturer's instructions using 120 ng total RNA as template and random hexamer primers. cDNA was diluted with DNase- and RNase-free water and amplified in a LightCycler® 480 II (Roche Diagnostics GmbH, Mannheim, Germany).

The *agrC*, *lukE*, *sdrC*, *srrA*, and *cap5A* genes were selected to verify the data obtained through RNA-seq. The gene *gyrB* was used as a control. These genes were amplified using TaqMan chemistry according to the instructions provided in LightCycler® 480 Probes Master Kit (Roche Diagnostics). All RT-qPCR reactions were carried out in a total volume of 20 µL. The amplification step was performed with 'Quantification' analysis mode at 95°C for 10 s, with a ramp rate of 4.4°C/s, followed by annealing at the respective annealing temperature for 50 s, with a ramp rate of 2.2°C/s and finally an extension at 72°C for 1 s, with a ramp rate of 4.4°C/s. The amplification step was performed 45 times. All primers and probes used in the study are listed in **Supplementary Table S1**. All RT-qPCR experiments were done in triplicate and each experiment was repeated at least twice. Data were analyzed by LightCycler® 480 Software release 1.5.0 by using the 'Relative Standard Curve' method; the standard curves were constructed using genomic DNA from *S. aureus* 04-02981. Data represent expression ratios, calculated by normalizing to the *gyrB* gene and relative to the untreated culture of *S. aureus* 04-02981 which served as the calibrator, as described in 'Guide to performing Relative Quantitation of Gene expression using real time-quantitative PCR' by Applied Biosystems. Means of five Cp values each were used to calculate the relative expression ratio.

## Statistical Analysis

Statistical tests were performed to analyze the significance of the obtained data. Student's *t*-test was applied to the normalized target, and normalized control values (normalized concentration). The tests were performed and analyzed using SigmaPlot version 11.0 (Systat software, Inc., San Jose, CA, United States<sup>3</sup>) (Wass, 2009).

## RESULTS

### AGXX® Inhibits the Growth of *S. aureus* 04-02981

To analyze the effect of Ag, and AGXX® on the growth of *S. aureus* 04-02981, disk diffusion tests with Ag, and AGXX® were performed in accordance with NCCLS-CLSI guidelines (Naas et al., 2006). The agar plates were monitored at 24 h intervals for

5 days to check if Ag or AGXX® exhibited an inhibitory effect on the pathogen, in the form of a zone of inhibition on the agar plate. The diameter of the inhibition zones was measured in 'cm.' The mean diameter of the inhibition zone was calculated to be 1.2 cm for AGXX® while no zone of inhibition was observed for Ag.

To verify the inhibitory effect of AGXX® on *S. aureus* 04-02981 as demonstrated in the agar diffusion tests, experiments in TSB medium were performed measuring the CFU/mL every hour for a period of 8 h, using the A: V ratio (metal mesh: medium volume) of 0.8, as described in Section "Materials and Methods." As observed in the disk diffusion assay, Ag did not show a significant inhibitory effect on the growth of *S. aureus* 04-02981 in liquid cultures. In contrast, AGXX® had a profound inhibitory effect on this strain. The OD<sub>600</sub> of *S. aureus* 04-02981 in presence of AGXX® was very low, (OD<sub>600</sub> AGXX® at t8 = 0.149) as compared to Ag (OD<sub>600</sub> Ag at t8 = 3.086) and the control (OD<sub>600</sub> Control at t8 = 3.173) (**Supplementary Table S2**). The CFU/mL of *S. aureus* 04-02981 grown in the batch culture with AGXX® increased from  $2.77 \times 10^6$  in the 1st hour to  $3.99 \times 10^{10}$  in the 4th hour, but then decreased to  $1.08 \times 10^7$  in the 8th hour. The colony counts of *S. aureus* 04-02981 + AGXX® (after 8 h of growth) were much lower than that of the same strain with Ag ( $1.27 \times 10^{11}$ ) or without metal amendment ( $1.73 \times 10^{11}$ ) (**Table 1**). These data confirm the antimicrobial effect of AGXX® on *S. aureus* 04-02981.

### AGXX® Strongly Reduces Biofilm Formation of *S. aureus* 04-02981

The effect of AGXX®, and Ag on biofilm formation of *S. aureus* 04-02981 was analyzed using the Crystal Violet assay. *E. faecalis* 12030, a strong biofilm former served as a positive control (Huebner et al., 1999), and TSB as the negative control (**Figure 1**). **Figure 1A** shows the biofilm formation by *S. aureus* 04-02981, measured at 570 nm, **Figure 1B** shows the biofilm formation (OD<sub>570</sub>) normalized to the bacterial growth (OD<sub>600</sub>) to take the antimicrobial effect of AGXX® into account.

To determine the bacterial killing activity of AGXX® under these conditions (after 24 h of growth, prior to adding crystal violet), we measured the CFU per mL of the planktonic cultures and the biofilms in the presence as well as in absence of the two different metal sheets.

The following values were obtained for the biofilms: For *S. aureus* 04-02981 without metal sheet (control),  $2.34 \times 10^9 \pm 8.49 \times 10^7$  CFU per mL, for the strain in presence of Ag,  $2.13 \times 10^9 \pm 2.40 \times 10^8$ , and in presence of AGXX®,  $1.80 \times 10^4 \pm 1.41 \times 10^3$ . When we measured the CFU per mL in the respective planktonic cultures, for the control,  $2.55 \times 10^8 \pm 2.12 \times 10^7$ , and for the strain in presence of Ag,  $2.00 \times 10^8 \pm 1.41 \times 10^7$  CFU per mL were obtained. However, no colonies were observed in presence of AGXX®. Thus, we conclude that in contrast to Ag, all planktonic bacteria were killed by AGXX® and after exposure to AGXX®, only a drastically reduced number of bacteria ( $1.80 \times 10^4$  CFU per mL) survived in the biofilm in comparison to Ag ( $2.13 \times 10^9$  CFU per mL).

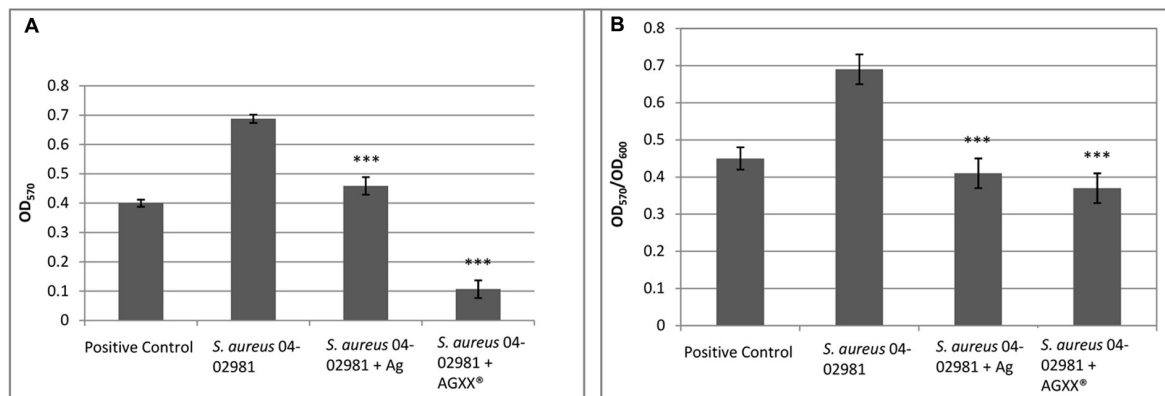
In summary, the biofilm formation measures normalized to the bacterial growth show that AGXX® reduced biofilm

<sup>3</sup> www.systatsoftware.com

**TABLE 1** | Colony forming units (CFU)/mL of *Staphylococcus aureus* 04-02981 (without sheet = control), in the presence of AGXX® or Ag.

Sample	0 h	1 h	2 h
Control	$1.27 \times 10^6 \pm 2.0 \times 10^5$	$7.63 \times 10^7 \pm 1.5 \times 10^6$	$8.71 \times 10^9 \pm 1.0 \times 10^9$
Ag	$1.23 \times 10^6 \pm 4.7 \times 10^5$	$6.80 \times 10^6 \pm 1.1 \times 10^6$	$6.23 \times 10^9 \pm 1.2 \times 10^9$
AGXX®	$7.67 \times 10^5 \pm 4.1 \times 10^5$	$2.77 \times 10^6 \pm 1.4 \times 10^6$	$4.93 \times 10^7 \pm 3.4 \times 10^7$
	3 h	4 h	5 h
Control	$1.11 \times 10^{10} \pm 1.1 \times 10^9$	$1.26 \times 10^{11} \pm 7.5 \times 10^9$	<b><math>1.73 \times 10^{11} \pm 3.3 \times 10^9</math></b>
Ag	$1.02 \times 10^{10} \pm 1.2 \times 10^9$	$1.25 \times 10^{11} \pm 1.0 \times 10^{10}$	<b><math>1.27 \times 10^{11} \pm 6.2 \times 10^8</math></b>
AGXX®	$2.06 \times 10^8 \pm 4.7 \times 10^7$	$3.99 \times 10^{10} \pm 1.1 \times 10^9$	<b><math>1.71 \times 10^8 \pm 1.6 \times 10^7</math></b>
	6 h	7 h	8 h
Control	$1.66 \times 10^{11} \pm 5.7 \times 10^9$	$1.26 \times 10^{11} \pm 3.5 \times 10^9$	<b><math>1.17 \times 10^{11} \pm 4.4 \times 10^9</math></b>
Ag	$1.22 \times 10^{10} \pm 4.6 \times 10^8$	$1.12 \times 10^{10} \pm 1.4 \times 10^9$	<b><math>1.02 \times 10^{10} \pm 1.4 \times 10^9</math></b>
AGXX®	$1.99 \times 10^8 \pm 7.1 \times 10^8$	$1.34 \times 10^7 \pm 1.4 \times 10^6$	<b><math>1.08 \times 10^7 \pm 1.8 \times 10^6</math></b>

The values for 5th hour and 8th hour are bolded because after  $t = 5h$ , the CFU values of MRSA + Ag decreased. And until  $t = 8h$ , the CFU values for all the three samples (MRSA, MRSA + Ag, and MRSA + AGXX) decreased.



**FIGURE 1** | AGXX® reduces biofilm formation of *S. aureus* 04-02981. Mean of five OD values of the biofilm assay with AGXX®, Ag, and without metal amendment with standard deviation are shown. **(A)** Shows the biofilm formation by *Staphylococcus aureus* 04-02981, measured at 570 nm; **(B)** shows the biofilm formation (OD<sub>570</sub>) normalized to the bacterial growth (OD<sub>600</sub>) to take the antimicrobial effect of AGXX® into account. Asterisks indicate the  $p$ -values obtained from  $t$ -test using SigmaPlot (\*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ ) representing the statistical significance of the data.

formation of *S. aureus* 04-02981 by 46%, whereas the inhibitory effect of Ag on biofilm formation was less pronounced (41%).

The strong reduction of biofilm formation by AGXX® was confirmed by Hoechst 33342/propidium iodide staining of biofilms grown for 24 h in presence of AGXX®, Ag and without antimicrobial sheet (Figure 2). The inhibitory effect of Ag was also clearly visible, although it was less distinct.

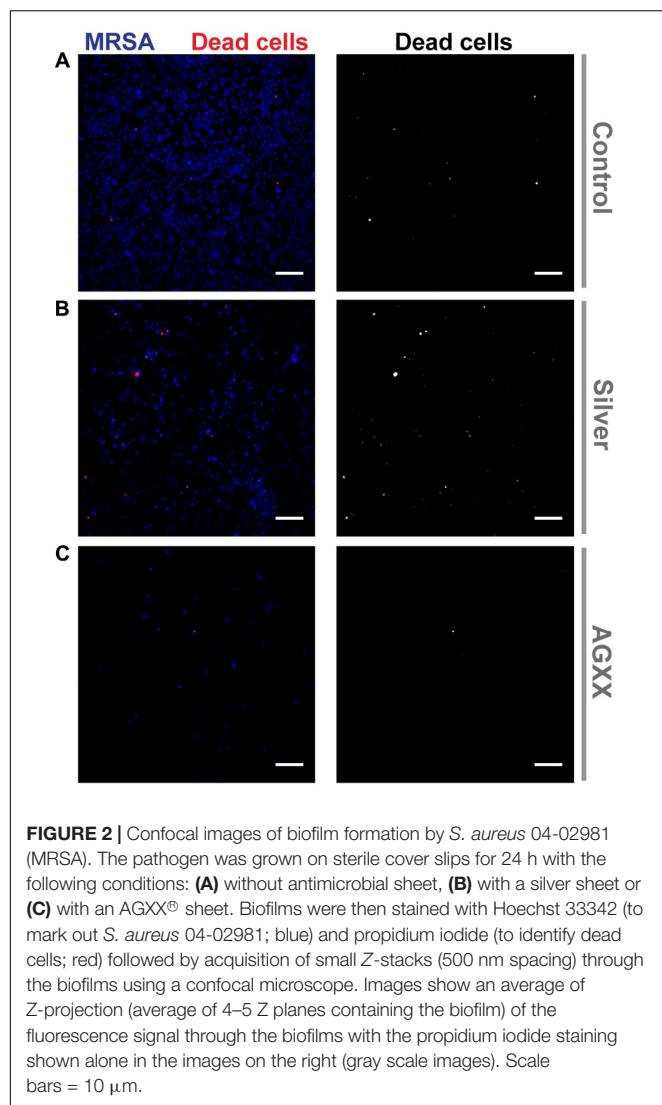
### AGXX® Strongly Induces Stress Response and Represses Pathogenesis in *S. aureus* 04-02981

The raw RNA sequence data obtained were aligned to the *S. aureus* 04-02981 genome. High sequencing depth was achieved as a mean value of ~12.4 million reads was obtained. The numbers of reads per sample ranged from ~8.4 million reads (Ag\_24) to 175 million reads (Control\_120) (Supplementary Table S3 and Supplementary Figure S1). From the data, it

is clear that the antimicrobial coating has a strong impact on the transcriptome of *S. aureus* 04-02981. In total, 2864 genes were differentially expressed in *S. aureus* 04-02981 on exposure to AGXX® and Ag (Supplementary Table S4). The number of differentially expressed genes in presence of AGXX® or Ag at different time-points is presented in Figure 3.

From Figure 3A, it can be seen that the number of differentially expressed genes at t24, t80, and t120 was quite similar. The maximum impact of AGXX® on the transcriptome of *S. aureus* 04-02981 was reached already after exposure for 24 min (723 genes up-regulated and 823 genes down-regulated) and remained nearly the same after exposure for 80 min (716 genes up- and 822 genes down-regulated), and 120 min (726 genes up- and 836 genes down-regulated). The lowest number of genes was differentially expressed at t6.

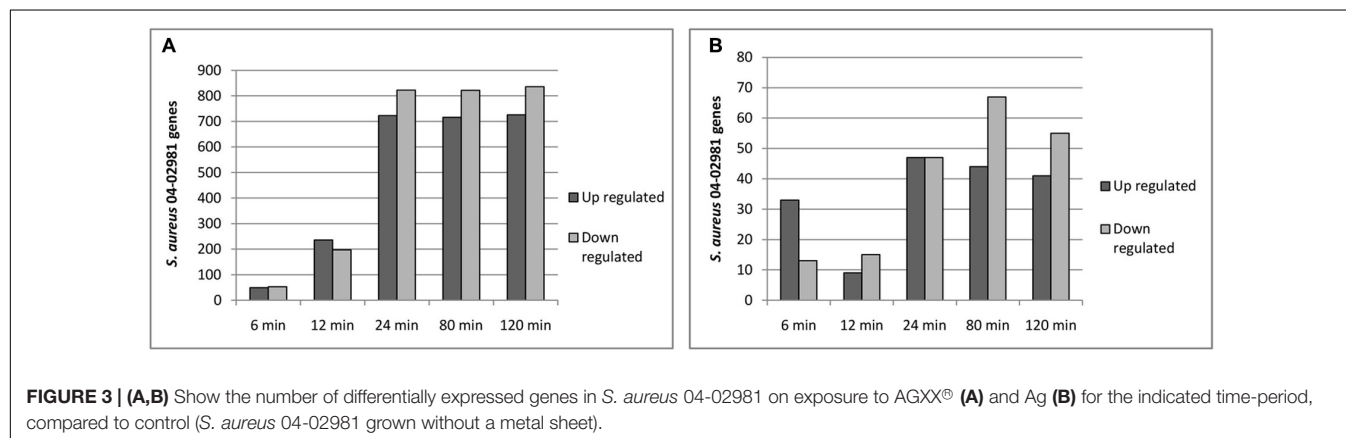
The differentially expressed genes were categorized as per Gene Ontology (GO) using the GSEA\_Pro option in the RNA-Seq analysis section in the T-REx RNA-Seq analysis pipeline (de Jong et al., 2015). Several GOs were obtained via

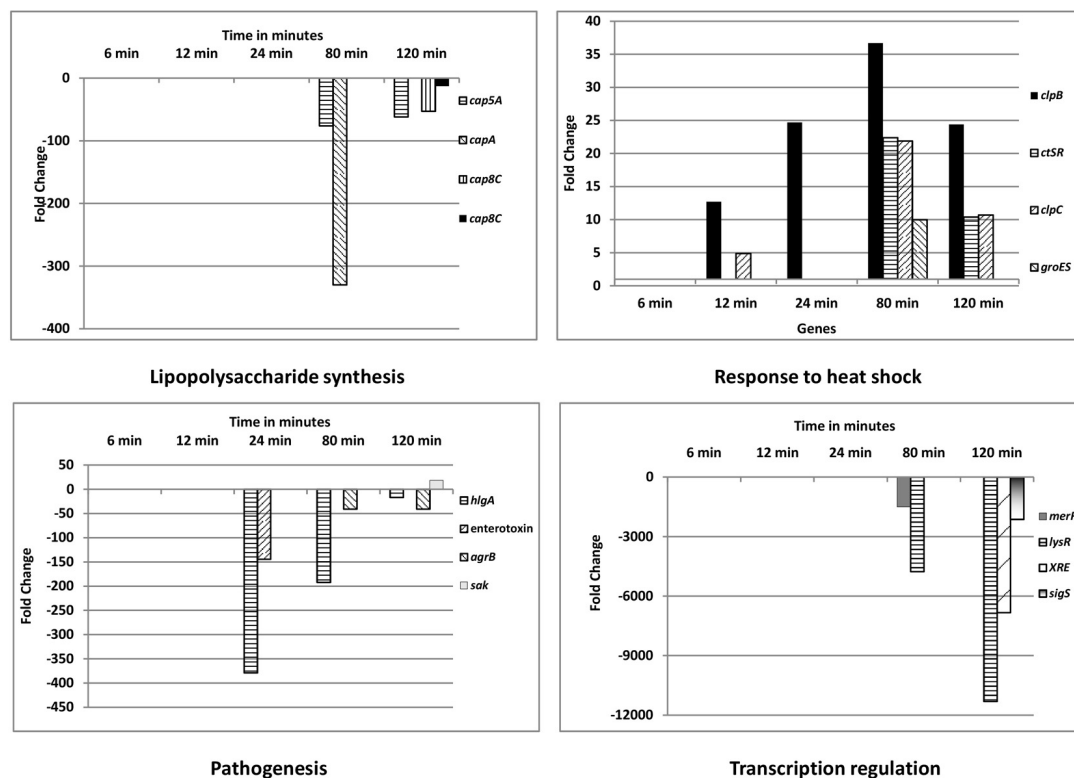


GSEA\_Pro, namely, oxidoreductase process, lipopolysaccharide synthesis, ATP binding, membrane transport, metabolism, metal binding, pathogenesis, transcription regulation, response

to heat shock, iron-siderophore transporter activity, serine protease activity, etc. (**Supplementary Table S5**). In the GO “lipopolysaccharide synthesis,” the *cap* genes mediating capsular polysaccharide synthesis (*cap5A*, *capA*, and *cap8C*) were all down-regulated. Genes (*clpB*, *ctsR*, *clpC*, and *groES*) involved in response to heat shock were up-regulated. Among the genes related to virulence (pathogenesis), 10 out of 11 genes were down-regulated, while only one gene was up-regulated at t120 (staphylokinase, a plasminogen activator). Among the responding transcriptional regulator genes, nine were up-regulated and 25 were down-regulated. **Figure 4** shows the differential expression of these GOs in *S. aureus* 04-02981 exposed to AGXX®.

*hlgA* (SA2981\_RS09385) was the most differentially expressed gene associated with virulence, it was down-regulated at t24 (378 fold), at t80 (192 fold), and at t120 (16 fold). The protein encoded by *hlgA* functions as a two-component toxin along with leukocidins in the lysis of erythrocytes (Gouaux et al., 1997). Among the transcriptional regulators, the gene of the LysR family transcriptional regulator, *lysR* was the most significantly influenced one by AGXX®, being down-regulated about 4700 fold at t80, and about 11,000 fold at t120. One of the LysR family transcriptional regulators, HutR is involved in metabolic processes of *S. aureus* (Ibarra et al., 2013). AGXX® had the highest impact on the expression of *capA*, of all the genes mediating capsular polysaccharide synthesis. *capA* was down-regulated by 329 fold at t80. Among the most differentially expressed genes in response to heat shock was *clpB*. It is a member of the stress-induced multi-chaperone system and works with DnaK, DnaJ, and GrpE in the recovery of the cell from heat-shock damage (Frees et al., 2005). Among the genes in the GO families influenced by AGXX®, only those involved in enterotoxin (SA2981\_RS09440), and staphylokinase production were also influenced by Ag, by -533 fold, and -2 fold, respectively, at t80 (**Supplementary Table S6**). In addition to the GO families, the effect of AGXX®, and Ag on the expression of operons in the pathogen was analyzed using the GSEA\_Pro option on the T-Rex pipeline. The results are presented in **Supplementary Tables S7, S8**, respectively.





**FIGURE 4 |** Gene Ontology (GO) categories of differentially expressed genes of *S. aureus* 04-02981 exposed to AGXX® for different time-periods. Only significantly differentially expressed genes, which are likely relevant in the response of *S. aureus* to AGXX®, are presented. Details on the other genes are provided in the **Supplementary Table S5**, along with all of the other GOs, Gene IDs, and fold-change in gene expression, for each time-point. Capsular polysaccharide synthesis mediating *cap* genes are shown in 'Lipopolysaccharide synthesis.' Genes encoding chaperone ClpB (*clpB*), transcriptional regulator CtsR (*ctsR*), ATP-dependent Clp protease ATP-binding subunit ClpC (*clpC*), and chaperonin GroES (*groES*) are presented in the panel showing response to heat shock. In the pathogenesis panel, the genes for  $\gamma$ -hemolysin subunit A (*hlgA*), enterotoxin, accessory gene regulator subunit B (*agrB*), and staphylokinase (*sak*) are presented. Among the transcriptional regulator genes, that of the LysR family transcriptional regulator (*lysR*), MerR family transcriptional regulator (*merR*), XRE family transcriptional regulator (*XRE*) and RNA polymerase sigma factor SigS (*sigS*) are shown.

## AGXX® Represses the Expression of Biofilm and Virulence-Associated Genes

We checked the effect of AGXX®, and Ag on the expression of genes associated with biofilm formation and virulence in *S. aureus* 04-02981. Many genes that are known to be crucial for biofilm formation and virulence were differentially expressed on exposure to AGXX® while Ag had an effect on just a few of them. The genes affected by AGXX® encode virulence factors, methicillin resistance, surface adhesins, capsular polysaccharide, two-component systems, and other biofilm-associated genes, as well as toxins (Table 2).

Upon exposure to AGXX®, the QS system genes *agrA*, *agrB*, *agrC*, and *agrD* of *S. aureus* 04-02981 were all down-regulated. Genes involved in the synthesis of capsular polysaccharide were also down-regulated. In general, the response of *S. aureus* 04-02981 to AGXX® was clearly visible after 24 min of exposure time. Genes encoding adhesins, *isdC*, *srtB*, and *sdrC* were also down-regulated. The *mecA* gene was down-regulated at t24. The up-regulation of genes inducing biofilm formation in *S. aureus*, such as *saeR* (2.3 fold at t120), *icaA* (36 fold at t24, 29 fold at t80 and 27 fold at t120), *icaB* (8 fold at t120) and *icaD* (55 fold at t12,

and 6 fold at t120) was intriguing. The genes *icaB*, *icaA*, and *icaD* are involved in *ica*-dependent biofilm formation. In addition, other key genes associated with biofilm formation and virulence, such as, *codY*, *srrA*, *luxS*, and genes for toxins like leukocidins, enterotoxins, hemolysins, were all differentially expressed at least at one of the time-points (Figure 5). Description of all locus tags and Gene IDs shown to the right of the heatmap is given in Table 2.

In general, it was observed that AGXX® had a huge impact on the transcriptome of *S. aureus* 04-02981, in particular at the later time-points 24, 80, and 120 min. In contrast, the effect of Ag was much less pronounced as already visible in the growth kinetics and to a lesser extent in the biofilm assays. Although, quite a number of *S. aureus* 04-02981 genes were differentially expressed upon exposure to Ag, only very few belong to the group of biofilm or virulence-associated genes. Among those, which were significantly differentially expressed in the presence of Ag, were *fmtC*, which is associated with methicillin resistance (approximately 3 fold up-regulated at t80; in the presence of AGXX® it was 2 fold up-regulated at t24), transcriptional regulator *sarR* (approximately 3 fold down-regulated at t24;



**TABLE 2 |** Differential expression of biofilm, and virulence-associated genes in *S. aureus* 04-02981 on exposure to AGXX®.

Locus tag	Abbreviation	Description	6 min	12 min	24 min	80 min	120 min
SA2981_RS10640	<i>agrD</i>	Accessory gene regulator D			−8.9	−11.9	−17.5
SA2981_RS10645	* <i>agrC</i>	Histidine kinase of the competence regulon ComD			−9.9	−7.7	−7
SA2981_RS10635	<i>agrB</i>	Accessory gene regulator B		−5.7	−18.3	−40.7	−40.8
SA2981_RS10650	<i>agrA</i>	Two-component system, LytR family, response regulator AgrA			−2.2		
SA2981_RS05970	PSM- $\beta$	Phenol-soluble modulins Beta				−10.5	−22.7
SA2981_RS05965	PSM- $\beta$	Phenol-soluble modulins Beta				−10.2	−12.4
SA2981_RS10825	<i>sigB</i>	RNA polymerase Sigma-B factor			−2.3	−4	−4.9
SA2981_RS07680	* <i>srrA</i>	DNA-binding response regulator SrrA		−4.6	−9.1	−4.6	−5.3
SA2981_RS00190	<i>mecA</i>	mecA-Penicillin- binding Protein 2		−2.3	−5.5		
SA2981_RS12040	<i>sarR</i>	Transcriptional regulator SarR					
SA2981_RS06390	<i>codY</i>	GTP-sensing transcriptional pleiotropic repressor CodY			−2.3	2.4	3.2
SA2981_RS00550	<i>sarH1</i>	Staphylococcal accessory regulator A			−2.4		
SA2981_RS05325	<i>FmtA</i>	FmtA protein involved in methicillin resistance				−2.6	
SA2981_RS05940	<i>arcD</i>	Arginine/ornithine antiporter ArcD					2.4
SA2981_RS00770	<i>capF</i>	Capsular polysaccharide synthesis enzyme Cap8F			−4.8	−3.1	−2
SA2981_RS05275	<i>sspb</i>	Staphopain B precursor				−2.3	
SA2981_RS13390	<i>cidA</i>	Holin-like protein				2.8	8.2
SA2981_RS13925	<i>arcA</i>	Arginine deiminase	−2.6	−11.8	−114.7	−20	−4.6
SA2981_RS03620	<i>saeR</i>	two-component system, OmpR family, response regulator SaeR					2.3
SA2981_RS06960	<i>FmtC</i>	Protein involved in methicillin resistance/L-lysine modification of phosphatidylglycerol			2.3		
SA2981_RS05900	hemolysin II	Alpha-hemolysin precursor				3.8	3.7
SA2981_RS01335	<i>IrgA</i>	Antiholin-like protein		−2.6	−3.1		−3.7
SA2981_RS00745	* <i>capA</i>	Capsular polysaccharide synthesis enzyme Cap5A			−34.9	−76.9	−61.9
SA2981_RS00750	<i>capB</i>	Tyrosine-protein kinase EpsD/capsular polysaccharide synthesis enzyme		2.4	−4.9	−29	−38.1
SA2981_RS00755	<i>capC</i>	protein-tyrosine phosphatase/capsular polysaccharide synthesis enzyme		2.8	−15	−20205	−52.9
SA2981_RS13940	<i>aur</i>	Zinc metalloproteinase precursor/aureolysin			−12.4		
SA2981_RS02875	* <i>sdrC</i>	Serine-aspartate repeat-containing protein C			−13.4		−9.8
SA2981_RS02035	Exotoxin 6	Superantigen-like protein			−8.7	4	−12.6
SA2981_RS13920	<i>arcB</i>	Ornithine carbamoyltransferase			−16.3	−26.1	−8.6
SA2981_RS05715	<i>isdC</i>	NPQTN cell wall anchored protein IsdC				−5	−9.1
SA2981_RS05735	<i>srtB</i>	Sortase B				−8.3	−22.4
SA2981_RS14090	<i>icaD</i>	Polysaccharide intercellular adhesin (PIA) biosynthesis protein		54.7			5.6
SA2981_RS14085	<i>icaA</i>	Polysaccharide intercellular adhesin (PIA) biosynthesis N-glycosyltransferase			35.7	29.9	26.5
SA2981_RS14095	<i>icaB</i>	Polysaccharide intercellular adhesin (PIA) biosynthesis deacetylase				−102.4	7.8
SA2981_RS09385	* <i>lukE</i>	Leukotoxin/leukocidin			−378.9	−192.2	

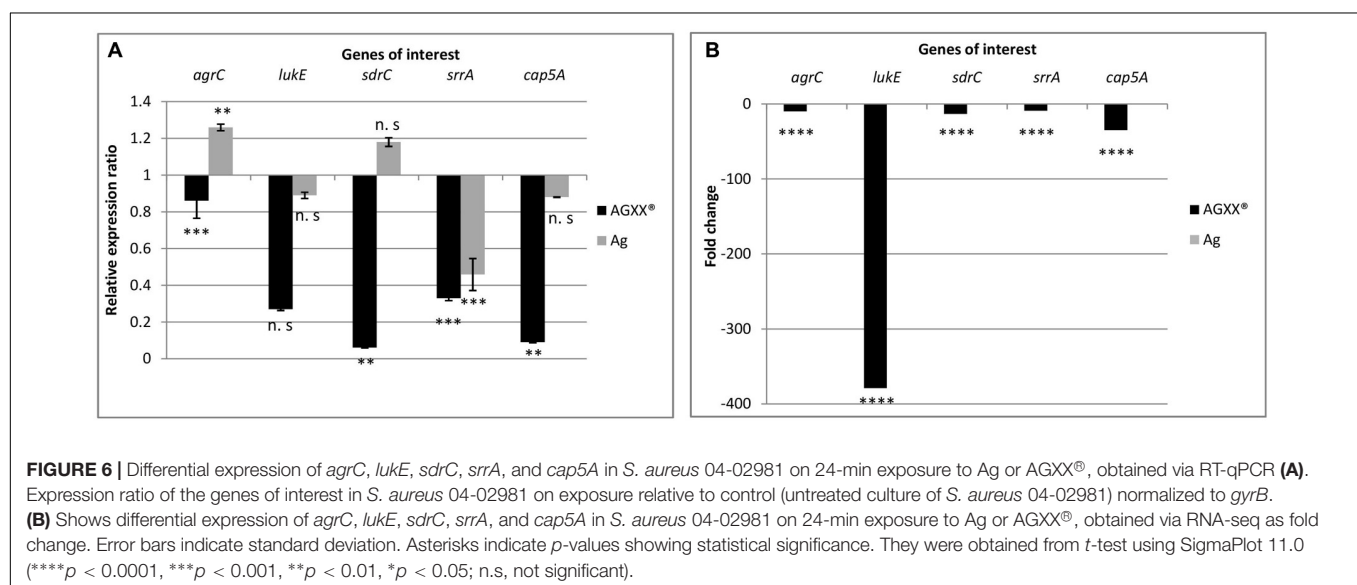
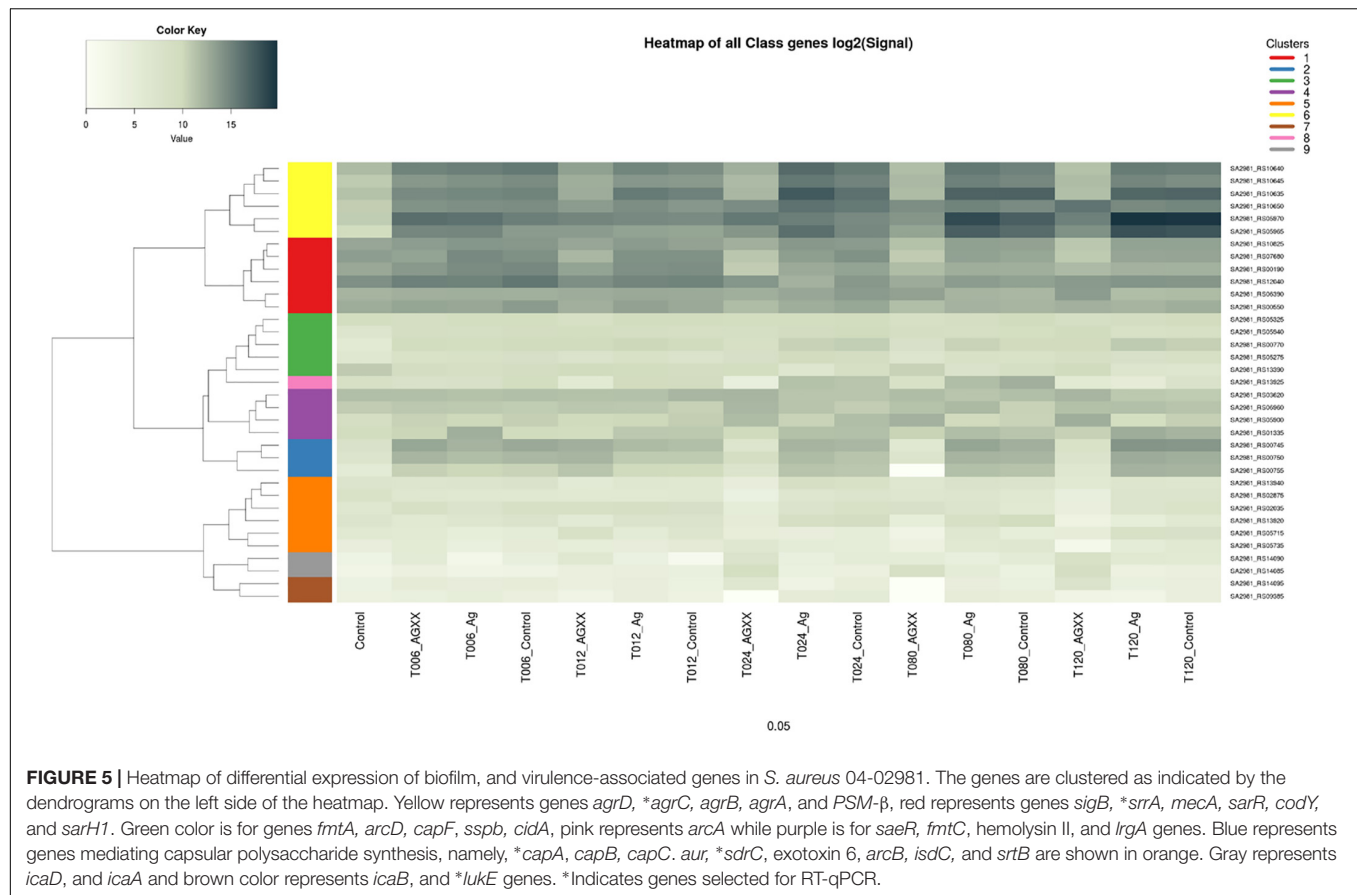
\*Genes selected for validation via RT-qPCR.

not differentially expressed in the presence of AGXX®, the gene of the holin-like protein CidA (approximately 4 fold down-regulated at t24; ~2 and ~8 fold up-regulated at t80 and t120, respectively, with AGXX®), the arginine deaminase gene *arcA* (approximately 6 fold down-regulated at t120 and 4.6 fold down-regulated with AGXX®), the hemolysin II gene (approximately 2 fold down-regulated at t24 and approximately 3 fold down-regulated at t120; ~3.7 fold up-regulated with AGXX® at t80 and t120) and the gene of the antiholin-like protein IrgA

(approximately 6 fold up-regulated at t6 with Ag, in the presence of AGXX®, it was ~3- to 3.7 fold down-regulated at t24, t80, and t120).

## Validation of RNA-Sequencing Data Using RT-qPCR

From the RNA-seq data, we observed that AGXX® affected genes encoding two-component systems, surface adhesins, capsular polysaccharides, and toxins. In total, five, highly



differentially expressed genes encoding these functions were selected to validate the RNA-seq derived transcriptional response of *S. aureus* 04-02981 to exposure to Ag or AGXX®. The validation experiment was performed on RNA extracted from *S. aureus* 04-02981 cultures exposed for 24, and 80 min to Ag or AGXX® since the selected genes were most differentially

expressed at these time-points. The five selected genes were, *agrC*, and *srrA* which are part of the two-component systems AgrCA and SrrAB, respectively (Baker et al., 2010; Wu et al., 2015), *lukE* which encodes a toxin (Liu et al., 2016), *sdrC* specifying a surface adhesin (Barbu et al., 2014), and *cap5A* mediating the synthesis of capsular polysaccharides

(Qin et al., 2014). *gyrB* was used as the house-keeping gene (Smith et al., 2010; Cheung et al., 2011). **Figures 6, 7** show the results of these experiments.

After exposure to AGXX® for 24 min, all five genes were down-regulated both in RNA-seq analysis and in RT-qPCR studies as can be seen in **Table 2**, and **Figure 6**. However, after exposure to AGXX® for 80 min, *sdrC* was down-regulated in RT-qPCR assays but it was not differentially expressed in RNA-seq. All the other genes were down-regulated in both approaches as seen in **Table 2** and **Figure 7**, respectively. On exposure to AGXX® for 24 min, *sdrC* was the most down-regulated gene followed by *cap5A*, *lukE*, *srrA*, and *agrC*, whereas after 80 min, *agrC* was the most down-regulated gene followed by *srrA*, *lukE*, *cap5A*, and *sdrC*. On exposure to Ag for 24 min, *srrA* was the most down-regulated gene, whereas *agrC* was the most up-regulated gene, and after 80 min, *cap5A* was the most down-regulated gene while *sdrC* was the only up-regulated gene, as observed in the RT-qPCR experiments.

## DISCUSSION

Multiple drug resistant, biofilm forming nosocomial pathogens such as MRSA pose a severe threat to public health demanding the development of novel antimicrobials as well as potent biofilm inhibitors. AGXX® is an effective antimicrobial that is active against many Gram-positive and Gram-negative bacteria (Guridi et al., 2015). AGXX® has been demonstrated to kill *S. aureus* 04-02981 as shown here by disk diffusion assay and growth kinetics experiments. In addition, AGXX® inhibited biofilm formation of *S. aureus* 04-02981 by ~46%. Moreover, for all time-points examined, the number of differentially expressed *S. aureus* 04-02981 genes was much higher upon exposure to AGXX® (in total 2391) than to Ag (317). For t120, the time-point showing the highest number of differentially expressed *S. aureus* 04-02981 genes, 1562 genes were differentially expressed in presence of AGXX®, while only 96 genes were affected by Ag.

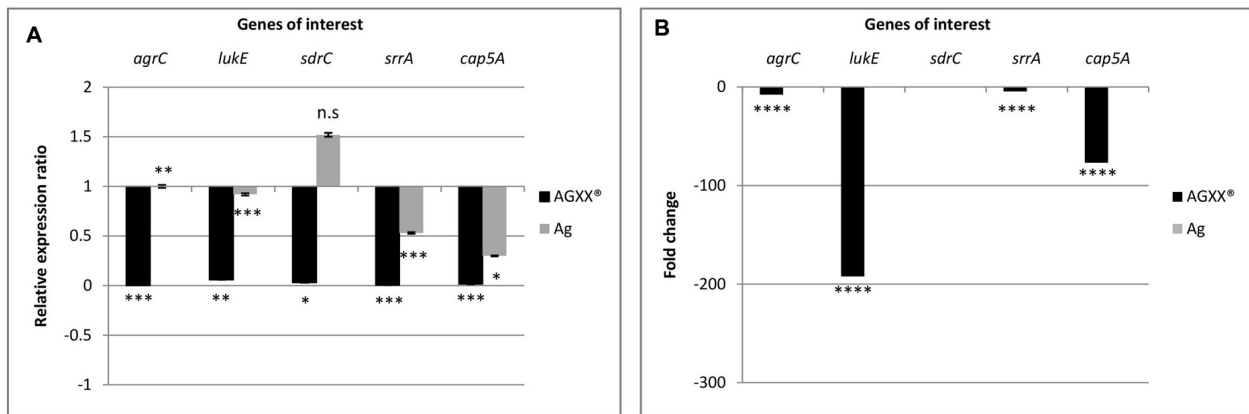
Up-regulation of genes of Gene Ontology (GOs) groups “response to heat shock” and “oxidoreductases” involved in oxidative stress response, and down-regulation of genes of GOs “pathogenesis” and “lipopolysaccharide synthesis” involving genes mediating capsular polysaccharide synthesis important for biofilm formation, point to a role of AGXX® as an antimicrobial and potent biofilm inhibitor. Together with results of a recent study where we have shown that the QS system of *S. aureus* 04-02981, *agr* was completely repressed after 4 h of exposure to AGXX® (Probst et al., 2016), we propose that AGXX® acts as a potential biofilm inhibitor. In *S. aureus*, two main mechanisms of biofilm formation are known, namely *ica*-dependent biofilm formation, which involves the production of polysaccharide intercellular adhesin (PIA), and *ica*-independent biofilm formation (Kirmusaoglu, 2016). Here we show that, in the presence of AGXX®, *icaA*, *icaD* were up-regulated and *icaB* was down-regulated. *icaA* and *icaD* contribute to the production of PIA (polymer). *icaD* transfers PIA to the cell surface of the bacteria while *icaB* deacylates PIA by fixing PIA to the outer surface of the

bacteria (Kirmusaoglu, 2016). In our study, intercellular adhesion biosynthesis *N*-deacetylase, *icaB* gene was down-regulated at t80 by ~100 fold. The structural development of exopolysaccharide-based biofilm requires deacetylation of PIA (Arciola et al., 2015). Since *icaB* was strongly down-regulated at t80, deacetylation of PIA probably does not occur which would obstruct the development of an exopolysaccharide-based biofilm. Fitzpatrick et al. (2005) showed that biofilm formation was unaffected in an *icaADBC* operon-deleted MRSA strain, while the same mutation in a methicillin sensitive strain of *S. aureus* (MSSA) impaired biofilm formation, suggesting strain-specificity in *ica*-dependent biofilm formation.

A two-component system associated with *ica*-dependent biofilm formation is *SrrAB* that acts as an autoregulator of biofilm formation. Deletion of *srrAB* inhibited *S. aureus* biofilm formation under oxic as well as microaerobic conditions (Wu et al., 2015). In our study, *srrA* was down-regulated 4 to 5 fold after 24, 80, and 120 min of exposure to AGXX®.

Global regulatory systems such as the *agr* QS system are among the best-studied factors involved in *ica*-(PIA) independent biofilm formation. Other proteins involved in such biofilms are SasG, SasC, Protein A, FnbB, FnbA, ATLA or ATLE, SdrG, SdrC, SdrD, biofilm associated protein (Bap) and lipoteichoic acid (Kirmusaoglu, 2016). We observed that two of these genes were down-regulated when AGXX® was present, namely *sdrD* and *sdrC*, *sdrC* was down-regulated 13- to 10 fold at t24 and t120, while *sdrD* was down-regulated 2 to 3 fold at t24 and t80. Moreover, the expression of lipoteichoic acid synthase, an enzyme responsible for the synthesis of lipoteichoic acid (Karatsa-Dodgson et al., 2010) was down-regulated approximately 4 fold after 24, 80, or 120 min of AGXX® presence. These data suggest that AGXX® might be working in an *ica*-independent manner to inhibit biofilm formation.

The *agr* locus contains five genes, *agrA*, *agrB*, *agrC*, *agrD*, and *hld*. On exposing *S. aureus* 04-02981 to AGXX®, only *hld* was not differentially expressed at any time-point, while all the other four genes were significantly down-regulated. The *agr* gene cluster regulates the expression of virulence factors such as phenol soluble modulins (PSMs), proteins that are closely associated with human skin and soft tissue infections (SSTIs) (Sully et al., 2014). “AgrD is a precursor peptide of autoinducer peptide (AIP)” (Quave and Horswill, 2014), AgrB is a membrane protease, which is involved in proteolytic processing and export of AgrD. It is also involved in AIP production (Njoroge and Sperandio, 2009; Quave and Horswill, 2014). AgrBD produce and secrete AIPs. AgrC, a sensor histidine kinase is activated when AIPs bind to AgrC. As a consequence, AgrC undergoes phosphorylation to activate AgrA, which is a DNA-binding response regulator (Njoroge and Sperandio, 2009). In our study, the *agrB* gene was the most down-regulated, at t80, and t120 (approximately 41 fold in both cases), while *agrA* was differentially expressed only at t24 (2 fold down-regulated). At t12, only *agrB* was differentially expressed, approximately 6 fold down-regulated. None of the *agr* genes was differentially expressed at t6. PSMs are staphylococcal toxins playing a role in acute infection (Kirmusaoglu, 2016); they are required for maturation and detachment of biofilm (Ma



**FIGURE 7 |** Differential expression of *agrC*, *lukE*, *sdrC*, *srrA*, and *cap5A* in *S. aureus* 04-02981 on 80-min exposure to Ag or AGXX®, obtained via RT-qPCR. Expression ratio of the genes of interest in *S. aureus* 04-02981 on exposure relative to control (untreated culture of *S. aureus* 04-02981) normalized to *gyrB* are shown **(A)**. **(B)** Shows differential expression of *agrC*, *lukE*, *sdrC*, *srrA*, and *cap5A* in *S. aureus* 04-02981 on 80-min exposure to Ag or AGXX®, obtained via RNA-seq as fold change. Error bars indicate standard deviation. Asterisks indicate *p*-values showing statistical significance. They were obtained from *t*-test using SigmaPlot 11.0 (\*\*\*\**p* < 0.0001, \*\*\**p* < 0.001, \*\**p* < 0.01, \**p* < 0.05; n.s., not significant).

et al., 2012). PSMs were also down-regulated in presence of AGXX® by ~10 fold at t80, and by 12 and 23 fold at t120. *agr* also regulates the expression of *sspB* which encodes a cysteine protease. *sspB* is positively associated with biofilm formation (Ma et al., 2012). It was down-regulated by 2.3 fold at t80. Inactivation of the alternative sigma factor SigB decreases biofilm formation in *S. aureus* (Ma et al., 2012). In presence of AGXX®, *sigB* was down-regulated 2–5 fold at the longer exposure times (t24, t80, and t120). In summary, down-regulation of all of the genes mentioned in this paragraph will likely reduce biofilm formation by *S. aureus*.

The two component systems, AgrCA and SaeRS influence biofilm formation in *S. aureus*, by the production of PSMs and by suppressing the synthesis of extracellular proteases, respectively (Baldry et al., 2016). The extracellular proteases degrade proteins that are important for biofilm formation (Baldry et al., 2016). In *S. aureus*, the *saeRS* system regulates the production of many virulence factors such as leukocidins, superantigens, proteases, surface proteins, and hemolysins (Liu et al., 2016). The gene for LukE, which enables *S. aureus* evasion from phagocytic cells by damaging the phagocytes was strongly down-regulated at t24 (379 fold) and t80 (192 fold). SplA is a serine protease, which is directly controlled by the *saeRS* system. *splA* was down-regulated 135 fold after 80 min of AGXX® presence. Mutations in genes for extracellular proteases (*splABCDEF*) in *S. aureus* SH1000 induced an increase in extracellular protease activity, which was associated with a reduction in biofilm formation (Chen et al., 2013). These facts taken together with *saeRS* not being differentially expressed at any time-point in the presence of AGXX®, except for a slight 2.3 fold up-regulation of *saeR* at t120, might suggest that *saeR* is not expressed in the mid exponential phase of growth of *S. aureus* 04-02981.

Capsular polysaccharides are also possible targets of the *saeRS* system (Liu et al., 2016). They play an important role in the virulence of the organism (Tuchscher et al., 2010). The synthesis

of capsular polysaccharides is mediated by the *cap5ABCFG* genes (Qin et al., 2014). Among these genes, only *capG* was not differentially expressed, all other genes were significantly down-regulated, especially at t24, t80, and t120, suggesting a role of AGXX® in repression of virulence in *S. aureus* 04-02981.

Another QS system, which significantly influences biofilm formation and virulence in Staphylococci is the *luxS* system. *luxS* impacts biofilm formation in a similar way as *agr* does, but by regulating different factors. *luxS* negatively regulates biofilm formation via cell-cell interactions based on autoinducer 2 secretion (Xu et al., 2006). The gene was 2.9 fold up-regulated at t24 in the presence of AGXX®.

In addition, the genes *isdC*, *srtB*, *sdrC*, encoding adhesins, were all down-regulated in the pathogen exposed to AGXX®. Iron regulated surface determinant IsdC is necessary for the primary attachment of *S. aureus* to surfaces such as polystyrene, as well as for the accumulation phase of biofilm formation; as such, it induces biofilm formation (Missineo et al., 2014). IsdC is anchored to the cell wall by sortase B (Hammer and Skaar, 2011). Serine-aspartate repeat containing protein C precursor (*SdrC*) assists bacteria in adhering to surfaces and promotes biofilm formation (Barbu et al., 2014). In *S. aureus* 04-02981 exposed to AGXX®, *isdC* was down-regulated by 5 and 9 fold at t80 and t120, respectively. The sortase B gene *srtB* was also down-regulated in cells treated with AGXX®, at t80 (8 fold) and t120 (22 fold). *sdrC*, too, was down-regulated some 10 to 13 fold at t24 and t120. Thus, we suggest that AGXX® inhibits biofilm formation in *S. aureus* 04-02981, also by repressing the expression of adhesins.

Reverse transcription quantitative PCR assays were performed on RNA extracted from *S. aureus* 04-02981 cultures exposed to Ag or AGXX® for 24 min, and 80 min to validate the RNA-seq data. In RT-qPCR, on exposure to AGXX® for 24 min, *agrC*, *sdrC*, *srrA*, and *cap5A* were statistically significantly down-regulated, whereas the down-regulation of *lukE* was not statistically



significant. In agreement with these data, the five genes were also significantly down-regulated in RNA-seq. By contrast, none of the five genes was significantly differentially expressed after 24 min in presence of Ag, as determined by RNA-seq, whereas RT-qPCR revealed a statistically significant down-regulation of *srrA* and a statistically significant up-regulation of *agrC*. The difference in expression of the other three genes *lukE*, *sdrC*, and *cap5A* was statistically not significant. When *S. aureus* 04-02981 was exposed to AGXX® for 80 min, all the five genes were down-regulated in RT-qPCR. The effect was statistically significant while in RNA-seq all genes were significantly down-regulated except *sdrC*. On exposure to Ag for 80 min, only *sdrC* was non-statistically significantly up-regulated. Thus, the trends in gene expression of *S. aureus* 04-02981 on exposure to AGXX® observed in RNA-seq and in RT-qPCR were similar.

In previous studies by others, differential gene expression of *S. aureus* in planktonic and biofilm mode has been examined. Resch et al. (2005) observed that in biofilms, genes encoding polysaccharide intercellular adhesin, and enzymes associated with cell envelope synthesis were significantly up-regulated (Resch et al., 2005). To combat biofilms, many metals have been tested for their capacity to inhibit bacterial biofilm formation. Specifically, silver nanoparticles have received much attention with respect to their antimicrobial nature. However, the minimum concentration of silver nanoparticles (AgNPs) required to eliminate biofilm formation is considered to have toxic effects on mammalian cells (Loo et al., 2016). They studied the effect of AgNPs and curcumin nanoparticles (Cur-NPs) on *S. aureus* and discovered that the combination of both nanoparticles was more effective than the individual AgNPs or Cur-NPs. Curcumin interferes with the QS system as was observed by the down-regulation of genes involved in QS, upon exposure to the substance (Loo et al., 2016). Ma et al. (2012) investigated the effect of two novel anti-virulence compounds on growth and biofilm formation of *S. aureus*. The compounds inhibited biofilm formation by repressing genes associated with biofilm formation such as *lrgA*, *sdrD*, *sspB*, *sigB*, *codY*, which were also down-regulated in our studies at least at one of the five time-points (Ma et al., 2012).

In summary, based on our findings, we conclude that AGXX® is an effective antimicrobial substance which might also act as a biofilm inhibitor based on our molecular data. The mechanism of inhibition is likely *ica*-independent without the production of PIA, by interfering with the QS system and by repressing genes associated with surface adhesin and lipopolysaccharide synthesis. In addition, the antimicrobial might also reduce pathogenesis of *S. aureus* 04-02981 by down-regulating the synthesis of toxins and virulence factors.

## AUTHOR CONTRIBUTIONS

AV performed all the microbiological and molecular experiments, drafted the manuscript, and designed the figures. Adj supervised and discussed bioinformatics analyses of

RNA-seq, and prepared and deposited the RNA-seq data at NCBI. DW performed the confocal microscopy and analyzed the data. JK drafted part of the discussion and gave insightful suggestions on molecular biology of Gram-positive pathogens. EG designed the project and supervised all the experiments. All authors discussed and corrected the manuscript.

## FUNDING

This research was funded by DLR, German Aerospace Center (Grant No. 50WB1466 to EG).

## ACKNOWLEDGMENTS

We thank U. Landau and C. Meyer from Largentec GmbH, Berlin, for providing us with the antimicrobial AGXX® and for the helpful discussions and G. Werner and J. Bender from Robert Koch Institute, Wernigerode Branch, for the gift of *S. aureus* 04-02981.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00221/full#supplementary-material>

**FIGURE S1** | Library sizes of all the RNA samples (*S. aureus* 04-02981, *S. aureus* 04-02981 + Ag, and *S. aureus* 04-02981 + AGXX), at different time periods. The image indicates the read depth of each sample. The X-axis represents the experiment names as used in the factors file, and gene counts file during RNA-seq analysis via T-REx. The sample names comprise the metal sheet used, followed by the time of exposure. For example, sample 'AGXX\_06' represents *S. aureus* 04-02981 exposed to AGXX for 6 minutes. The Y-axis represents the total number of mapped reads.

**TABLE S1** | Primer and probe sequences used for RT-qPCR.

**TABLE S2** | AGXX®-mediated growth inhibition of *S. aureus* 04-02981 in batch cultures.

**TABLE S3** | Alignment rates of the RNA-sequences of *S. aureus* 04-02981.

**TABLE S4** | Differentially expressed genes in *S. aureus* 04-02981 on exposure to Ag, and AGXX®.

**TABLE S5** | Gene Ontology assignments on exposing *S. aureus* 04-02981 to AGXX® for 6 minutes. Rate = The rating values (1 to 5) reflect binned values based on:  $(\text{TopHits}/\text{ClassSize}) * -\log_2(\text{adj-pvalue})$ .

**TABLE S6** | Gene Ontology assignments on exposing *S. aureus* 04-02981 to Ag for 80 minutes. Rate = The rating values (1 to 5) reflect binned values based on:  $(\text{TopHits}/\text{ClassSize}) * -\log_2(\text{adj-pvalue})$ .

**TABLE S7** | Expression of operons in *S. aureus* 04-02981 on exposure to AGXX® for 6 minutes. Rate = The rating values (1 to 5) reflect binned values based on:  $(\text{TopHits}/\text{ClassSize}) * -\log_2(\text{adj-pvalue})$ .

**TABLE S8** | Expression of operons in *S. aureus* 04-02981 on exposure to Ag for 80 minutes. Rate = The rating values (1 to 5) reflect binned values based on:  $(\text{TopHits}/\text{ClassSize}) * -\log_2(\text{adj-pvalue})$ .

## REFERENCES

- Arcioli, C. R., Baldassarri, L., and Montanaro, L. (2001). Presence of *icaA* and *icaD* genes and slime production in a collection of staphylococcal strains from catheter-associated infections. *J. Clin. Microbiol.* 39, 2151–2156.
- Arciola, C. R., Campoccia, D., Ravaoli, S., and Montanaro, L. (2015). Polysaccharide intercellular adhesin in biofilm: structural and regulatory aspects. *Front. Cell. Infect. Microbiol.* 5:7. doi: 10.3389/fcimb.2015.00007
- Baker, J., Sithisak, S., Sengupta, S., Johnson, M., Jayaswal, R. K., and Morrissey, J. A. (2010). Copper stress induces a global stress response in *Staphylococcus aureus* and represses *sae* and *agr* expression and biofilm formation. *Appl. Environ. Microbiol.* 76, 150–160. doi: 10.1128/AEM.02268-09
- Baldry, M., Nielsen, A., Bojer, M. S., Zhao, Y., Friberg, C., Ifrah, D., et al. (2016). Norlichexanthone reduces virulence gene expression and biofilm formation in *Staphylococcus aureus*. *PLOS ONE* 11:e0168305. doi: 10.1371/journal.pone.0168305
- Barbu, E. M., Mackenzie, C., Foster, T. J., and Höök, M. (2014). SdrC induces staphylococcal biofilm formation through a hemophilic interaction. *Mol. Microbiol.* 94, 172–185. doi: 10.1111/mmi.12750
- Bouchard, A. (2011). AgXX Glass Microspheres. In Vitro Evaluation of Cytotoxicity by Neutral Red Assay Using MRC-5 Cell Line with a Direct Contact Procedure. Report 20100326STP. Dresden: APOGEPHA Arzneimittel GmbH.
- Chen, C., Krishnan, V., Macon, K., Manne, K., and Schneewind, O. (2013). Secreted proteases control autolysin-mediated biofilm growth of *Staphylococcus aureus*. *J. Biol. Chem.* 288, 29440–29452. doi: 10.1074/jbc.M113.502039
- Cheung, G. Y. C., Wang, R., Khan, B. A., Sturdevant, D. E., and Otto, M. (2011). Role of the accessory gene regulator *agr* in community-associated methicillin-resistant *Staphylococcus aureus* pathogenesis. *Infect. Immun.* 79, 1927–1935. doi: 10.1128/IAI.00046-11
- Chopra, I. (2007). The increasing use of silver-based products as antimicrobial agents: a useful development or a cause for concern? *J. Antimicrob. Chemother.* 59, 587–590. doi: 10.1093/jac/dkm006
- Clauss-Lendzian, E., Vaishampayan, A., de Jong, A., Landau, U., Meyer, C., Kok, J., et al. (2017). Stress response of a clinical *Enterococcus faecalis* isolate subjected to a novel antimicrobial surface coating. *Microbiol. Res.* doi: 10.1016/j.micres.2017.11.006
- de Jong, A., van der Meulen, S., Kuipers, O. P., and Kok, J. (2015). T-REx: transcriptome analysis webserver for RNA-seq expression data. *BMC Genomics* 16:663. doi: 10.1186/s12864-015-1834-4
- Ellis, M. W., Schlett, C. D., Millar, E. V., Crawford, K. B., Cui, T., Lanier, J. B., et al. (2014). Prevalence of nasal colonization and strain concordance in patients with community-associated *Staphylococcus aureus* skin and soft-tissue infections. *Infect. Control Hosp. Epidemiol.* 35, 1251–1256. doi: 10.1086/678060
- Fitzpatrick, F., Humphreys, H., and O’Gara, J. P. (2005). Evidence for *icaADBC*-independent biofilm development mechanism in methicillin-resistant *Staphylococcus aureus* clinical isolates. *J. Clin. Microbiol.* 43, 1973–1976.
- Frees, D., Chastanet, A., Qazi, S., Sorensen, K., Hill, P., Msadek, T., et al. (2005). Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in *Staphylococcus aureus*. *Mol. Microbiol.* 54, 1445–1462. doi: 10.1111/j.1365-2958.2004.04368.x
- Gouaux, A., Hobaugh, M., and Song, L. (1997).  $\alpha$ -hemolysin,  $\gamma$ -hemolysin, and leukocidin from *Staphylococcus aureus*: distant in sequence but similar in structure. *Protein Sci.* 6, 2631–2635. doi: 10.1002/pro.5560061216
- Grass, G., Rensing, C., and Solioz, M. (2011). Metallic copper as an antimicrobial surface. *Appl. Environ. Microbiol.* 77, 1541–1547. doi: 10.1128/AEM.02766-10
- Gupta, A., Matsui, K., Lo, J.-F., and Silver, S. (1999). Molecular basis for resistance to silver cations in *Salmonella*. *Nat. Med.* 5, 183–188. doi: 10.1038/5545
- Guridi, A., Diederich, A. K., Aguila-Arcos, S., Garcia-Moreno, M., Blasi, R., Broszat, M., et al. (2015). New antimicrobial contact catalyst killing antibiotic resistant clinical and water borne pathogens. *Mater. Sci. Eng. C Mater. Biol. Appl.* 50, 1–11. doi: 10.1016/j.msec.2015.01.080
- Hammer, N. D., and Skaar, E. P. (2011). Molecular mechanisms of *Staphylococcus aureus* iron acquisition. *Annu. Rev. Microbiol.* 65, 129–147. doi: 10.1146/annurev-micro-090110-102851
- Huebner, J., Wang, Y., Krueger, W. A., Madoff, L. C., Martirosian, G., Boisot, S., et al. (1999). Isolation and chemical characterization of a capsular polysaccharide antigen shared by clinical isolates of *Enterococcus faecalis* and vancomycin-resistant *Enterococcus faecium*. *Infect. Immun.* 67, 1213–1219.
- Ibarra, J. A., Pérez-Rueda, E., Carroll, R. K., and Shaw, L. N. (2013). Global analysis of transcriptional regulators in *Staphylococcus aureus*. *BMC Genomics* 14:126. doi: 10.1186/1471-2164-14-126
- Karatsa-Dodgson, M., Woermann, M. E., and Gruending, A. (2010). In vitro analysis of the *Staphylococcus aureus* lipoteichoic acid synthase enzyme using fluorescently labeled lipids. *J. Bacteriol.* 192, 5341–5349. doi: 10.1128/JB.00453-10
- Kirmusaoglu, S. (2016). “Staphylococcal biofilms: pathogenicity, mechanism and regulation of biofilm formation by Quorum-Sensing system and antibiotic resistance mechanisms of biofilm-embedded microorganisms,” in *Microbial Biofilms - Importance and Applications*, ed. D. Dhanasekaran (Rijeka: In Tech).
- Kluytmans, J., van Belkum, A., and Verbrugh, H. (1997). Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin. Microbiol. Rev.* 10, 505–520.
- Landau, U. (2013). AGXX - Eine nachhaltige Lösung für die Entkeimung wässriger Lösungen. *Galvanotechnik* 11, 2169–2184.
- Landau, U., Meyer, C., and Grohmann, E. (2017a). AGXX - Beitrag der Oberflächentechnik zur Vermeidung von Biofilmen (Teil 1). *Galvanotechnik* 108, 885–890.
- Landau, U., Meyer, C., and Grohmann, E. (2017b). AGXX - Beitrag der Oberflächentechnik zur Vermeidung von Biofilmen (Teil 2). *Galvanotechnik* 108, 1110–1121.
- Landsdown, A. B. G. (2010). A pharmacological and toxicological profile of silver as an antimicrobial agent in medical devices. *Adv. Pharmacol. Sci.* 2010:910686. doi: 10.1155/2010/910686
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lister, J. L., and Horswill, A. R. (2014). *Staphylococcus aureus* biofilms: recent developments in biofilm dispersal. *Front. Cell. Infect. Microbiol.* 4:178. doi: 10.3389/fcimb.2014.00178
- Liu, Q., Yo, W.-S., and Bae, T. (2016). The SaeRS two-component system of *Staphylococcus aureus*. *Genes* 7:81. doi: 10.3390/genes7100081
- Loo, C.-Y., Rohanizadeh, R., Young, P. M., Traini, D., Cavaliere, R., Whitchurch, C. B., et al. (2016). Combination of silver nanoparticles and curcumin nanoparticles for enhanced anti-biofilm activities. *J. Agric. Food Chem.* 64, 2513–2522. doi: 10.1021/acs.jafc.5b04559
- Ma, Y., Xu, Y., Yestrepky, B. D., Sorenson, R. J., Chen, M., Larsen, S. D., et al. (2012). Novel inhibitors of *Staphylococcus aureus* virulence gene expression and biofilm formation. *PLOS ONE* 7:e47255. doi: 10.1371/journal.pone.0047255
- Maillard, J., and Hartemann, P. (2013). Silver as an antimicrobial: facts and gap in knowledge. *Crit. Rev. Microbiol.* 39, 373–383. doi: 10.3109/1040841X.2012.713323
- Marathe, N. P., Nagarkar, S. S., Vaishampayan, A. A., Rasane, M. H., Samant, S. A., Dohe, V., et al. (2015). High prevalence of class 1 integrons in clinical isolates of methicillin-resistant *Staphylococcus aureus* from India. *Indian J. Med. Microbiol.* 33, 231–236. doi: 10.4103/0255-0857.154905
- Missineo, A., Poto, D. A., Geoghegan, J. A., Rindi, S., Heilbronner, S., Gianotti, V., et al. (2014). IsdC from *Staphylococcus lugdunensis* induces biofilm formation under low-iron growth conditions. *Infect. Immun.* 82, 2448–2459. doi: 10.1128/IAI.01542-14
- Naas, T., Coignard, B., Carbonne, A., Blanckaert, K., Bajolet, O., Bernet, C., et al. (2006). VEB-1 Extended-spectrum beta-lactamase-producing *Acinetobacter baumannii*, France. *Emerg. Infect. Dis.* 12, 1214–1222. doi: 10.3201/eid1208.051547
- Njoroge, J., and Sperandio, V. (2009). Jamming bacterial communication: new approaches for the treatment of infectious diseases. *EMBO Mol. Med.* 1, 201–210. doi: 10.1002/emmm.200900032
- Nuebel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S. K., et al. (2010). A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin resistant *Staphylococcus aureus*. *PLOS Pathog.* 6:e1000855. doi: 10.1371/journal.ppat.1000855

- Nyenje, M. E., Green, E., and Ndip, R. N. (2013). Evaluation of the effect of different growth media and temperature on the suitability of biofilm formation by *Enterobacter cloacae* strains isolated from food samples in South Africa. *Molecules* 18, 9582–9593. doi: 10.3390/molecules18089582
- Paniagua-Contreras, G., Sáinz-Espuñes, T., Monroy-Pérez, E., Rodríguez-Moctezuma, J. R., Arenas-Aranda, D., Negrete-Abascal, E., et al. (2012). Virulence markers in *Staphylococcus aureus* strains isolated from hemodialysis catheters of Mexican patients. *Adv. Microbiol.* 2, 476–487. doi: 10.4236/aim.2012.24061
- Paraje, M. G. (2011). "Antimicrobial resistance in biofilms," in *Science against Microbial Pathogens: Communicating Current Research and Technological Advances*, Vol. 2, ed. A. Mendez-Vilas (Badajoz: Formatex Research Center), 736–744.
- Probst, I., Vaishampayan, A., Kuechler, V., and Grohmann, E. (2016). Antimikrobielle Oberflächenbeschichtung tötet multiresistente Krankheitserreger. *Flug Reisemed.* 23, 14–17.
- Qin, N., Tan, X., Jiao, Y., Liu, L., Zhao, W., Yang, S., et al. (2014). RNA-Seq-based transcriptome analysis of methicillin-resistant *Staphylococcus aureus* biofilm inhibition by ursolic acid and resveratrol. *Sci. Rep.* 4:5467. doi: 10.1038/srep05467
- Quave, C. L., and Horswill, A. R. (2014). Flipping the switch: tools for detecting small molecule inhibitors of staphylococcal virulence. *Front. Microbiol.* 5:706. doi: 10.3389/fmicb.2014.00706
- Resch, A., Rosenstein, R., Nerz, C., and Goetz, F. (2005). Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions. *Appl. Environ. Microbiol.* 71, 2663–2676.
- Sause, W. E., Buckley, P. T., Strohl, W. R., Lynch, A. S., and Torres, V. J. (2015). Antibody-based biologics and their promise to combat *Staphylococcus aureus* infections. *Trends Pharmacol. Sci.* 37, 231–241. doi: 10.1016/j.tips.2015.11.008
- Schäberle, T. F., and Hack, I. M. (2014). Overcoming the current deadlock in antibiotic resistance. *Trends Microbiol.* 22, 165–167. doi: 10.1016/j.tim.2013.12.007
- Schiwon, K., Arends, K., Rogowski, K. M., Fuerch, S., Prescha, K., Sakinc, T., et al. (2013). Comparison of antibiotic resistance, biofilm formation and conjugative transfer of *Staphylococcus* and *Enterococcus* isolates from International Space Station and Antarctic research station Concordia. *Microb. Ecol.* 65, 638–651. doi: 10.1007/s00248-013-0193-4
- Smith, K., Gould, K. A., Gordon, R., Gemmell, C. G., Hinds, J., and Lang, S. (2010). Influence of tigecycline on expression of virulence factors in biofilm-associated cells of methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 54, 380–387. doi: 10.1128/AAC.00155-09
- Sully, E. K., Malachowa, N., Elmore, B. O., Alexander, S. M., Femling, J. K., Gray, B. M., et al. (2014). Selective chemical inhibition of *agr* Quorum Sensing in *Staphylococcus aureus* promotes host defense with minimal impact on resistance. *PLOS Pathog.* 10:e1004174. doi: 10.1371/journal.ppat.1004174
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tuchscherer, L., Loeffler, B., Buzzola, F. R., and Sordelli, D. O. (2010). *Staphylococcus aureus* adaptation to the host and persistence: role of loss of capsular polysaccharide expression. *Future Microbiol.* 5, 1823–1832. doi: 10.2217/fmb.10.147
- Warnes, S. L., and Keevil, C. W. (2013). Inactivation of norovirus on dry copper alloy surfaces. *PLOS ONE* 8:e75017. doi: 10.1371/journal.pone.0075017
- Wass, Y. A. (2009). SigmaPlot 11: Now with total sigmaStat integration. *Sci. Comput.* 26, 21.
- Wu, Y., Wu, Y., Zhu, T., Han, H., Liu, H., Xu, T., et al. (2015). *Staphylococcus epidermidis* SrrAB regulates bacterial growth and biofilm formation differently under oxic and microaerobic conditions. *J. Bacteriol.* 197, 459–476. doi: 10.1128/JB.02231-14
- Xu, L., Li, H., Vuong, C., Vadyvaloo, V., Wang, J., Yao, Y., et al. (2006). Role of the *luxS* Quorum-Sensing system in biofilm formation and virulence of *Staphylococcus epidermidis*. *Infect. Immun.* 74, 488–496.
- Yarwood, J. M., Bartels, D. J., Volper, E. M., and Greenberg, E. P. (2004). Quorum Sensing in *Staphylococcus aureus* biofilms. *J. Bacteriol.* 186, 1838–1850. doi: 10.1128/JB.186.6.1838-1850.2004

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Vaishampayan, de Jong, Wight, Kok and Grohmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# DNA Delivery and Genomic Integration into Mammalian Target Cells through Type IV A and B Secretion Systems of Human Pathogens

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### Reviewed by:

Elisabeth Grohmann,  
Beuth University of Applied Sciences,  
Germany

Jose Angel Ruiz-Masó,  
Centro de Investigaciones Biológicas  
(CSIC), Spain

### \*Correspondence:

Matxalen Llosa  
llosam@unican.es

### † Present address:

Coral González-Prieto,  
Department of Medicine, Division  
of Infectious Diseases, Massachusetts  
General Hospital – Department  
of Microbiology and Immunobiology,  
Harvard Medical School, Cambridge,  
MA, United States

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 June 2017

**Accepted:** 26 July 2017

**Published:** 22 August 2017

### Citation:

Guzmán-Herrador DL, Steiner S,  
Alperi A, González-Prieto C, Roy CR  
and Llosa M (2017) DNA Delivery  
and Genomic Integration into  
Mammalian Target Cells through Type  
IV A and B Secretion Systems  
of Human Pathogens.  
Front. Microbiol. 8:1503.  
doi: 10.3389/fmicb.2017.01503

Dolores L. Guzmán-Herrador<sup>1</sup>, Samuel Steiner<sup>2</sup>, Anabel Alperi<sup>1</sup>, Coral González-Prieto<sup>1†</sup>,  
Craig R. Roy<sup>2</sup> and Matxalen Llosa<sup>1\*</sup>

<sup>1</sup> Departamento de Biología Molecular, Universidad de Cantabria (UC), Instituto de Biomedicina y Biotecnología de Cantabria (IBBTec, UC-CSIC-SODERCAN), Santander, Spain, <sup>2</sup> Department of Microbial Pathogenesis, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, CT, United States

We explore the potential of bacterial secretion systems as tools for genomic modification of human cells. We previously showed that foreign DNA can be introduced into human cells through the Type IV A secretion system of the human pathogen *Bartonella henselae*. Moreover, the DNA is delivered covalently attached to the conjugative relaxase TrwC, which promotes its integration into the recipient genome. In this work, we report that this tool can be adapted to other target cells by using different relaxases and secretion systems. The promiscuous relaxase MobA from plasmid RSF1010 can be used to deliver DNA into human cells with higher efficiency than TrwC. MobA also promotes DNA integration, albeit at lower rates than TrwC. Notably, we report that DNA transfer to human cells can also take place through the Type IV secretion system of two intracellular human pathogens, *Legionella pneumophila* and *Coxiella burnetii*, which code for a distantly related Dot/Icm Type IV B secretion system. This suggests that DNA transfer could be an intrinsic ability of this family of secretion systems, expanding the range of target human cells. Further analysis of the DNA transfer process showed that recruitment of MobA by Dot/Icm was dependent on the IcmSW chaperone, which may explain the higher DNA transfer rates obtained. Finally, we observed that the presence of MobA negatively affected the intracellular replication of *C. burnetii*, suggesting an interference with Dot/Icm translocation of virulence factors.

**Keywords:** protein secretion, bacterial conjugation, *Legionella pneumophila*, *Coxiella burnetii*, *Bartonella henselae*, conjugative relaxase, intracellular pathogen, gene therapy

## INTRODUCTION

Bacterial Type IV secretion systems (T4SS) selectively deliver macromolecules to other cells or to the extracellular media. An outstanding feature of these secretion systems is their ability to secrete both, protein and DNA molecules, a particularity that distinguishes them from other types of secretion systems. In addition, the secreted substrates can be delivered to either prokaryotic



or eukaryotic cells. This plasticity allows T4SS to be involved in bacterial processes as diverse as horizontal DNA transfer or virulence (Christie, 2016).

Bacterial Type IV secretion systems are multiprotein complexes formed by different constitutive elements: a core complex spanning both bacterial membranes, which forms the transport conduit; a pilus-like appendage, whose function as a transport channel is still under debate; a series of cytoplasmic ATPases, which energize the transport process; and elements necessary to recruit and present the substrates to the translocation machine, including chaperones that are variable for each system (Zechner et al., 2012). Within the family of T4SS, two sub-families were described based on sequence homologies: The Type IV A-IV B secretion systems (T4ASS and T4BSS, respectively). The formers are homologous to the prototypical VirB T4SS of *Agrobacterium tumefaciens* and have been characterized extensively, both functionally and structurally (Chandran Darbari and Waksman, 2015). Members of this family form part of conjugative systems of plasmids such as R388 or RP4; others are encoded in the genomes of human pathogens such as *Bartonella henselae* (Bh), *Brucella melitensis* or *Helicobacter pylori* among others, and their main role is to inject virulence factors to the target human cell. Similarly, T4BSS members are encoded in conjugative plasmids such as F, and in the chromosomes of human pathogens such as *Legionella pneumophila* (Lp) and *Coxiella burnetii* (Cb). Research on T4BSS structure and function lags behind T4ASS; however, extensive work has been done regarding the role of T4BSS-delivered effectors within human cell (Hubber and Roy, 2010; Rolando and Buchrieser, 2014; Personnic et al., 2016).

As aforementioned, a distinctive feature of T4SS is their ability to secrete DNA molecules. This is the main molecular function of T4SS belonging to the conjugative machinery of self-transmissible plasmids (Cabezón et al., 2015). In order to secrete DNA, at least two components are essential in addition to the T4SS machinery: an origin of transfer (*oriT*), which is the DNA sequence required in *cis* on a DNA molecule to be transferred, and a conjugative relaxase, which cuts the DNA strand to be transferred at the *oriT*. Many plasmids also encode for accessory nicking proteins, which assist the DNA processing by the relaxase. The DNA is transferred as a single strand covalently attached to the relaxase, which itself is the substrate of the T4SS; the nucleoprotein complex enters the recipient cell, where the relaxase catalyzes the recircularization of the transferred DNA strand (Garcillán-Barcia et al., 2007; Gonzalez-Perez et al., 2007).

Notably, some conjugative relaxases have the ability to catalyze site-specific recombination between two copies of *oriT*. This phenomenon was first described for the R388 relaxase TrwC (Llosa et al., 1994). TrwC acts as a site-specific recombinase on supercoiled substrates containing minimal target sequences (Cesar et al., 2006). This ability is shared by some, but not all, conjugative relaxases, and it is unclear why. MobA, the relaxase of the mobilizable plasmid RSF1010 (virtually identical to plasmid R1162), is able to catalyze *oriT-oriT* recombination on single-stranded substrates but not on supercoiled plasmid substrates (Meyer, 1989). TrwC can also catalyze the integration

of the transferred DNA molecule into a target sequence present in the recipient bacterium (Draper et al., 2005); moreover, the protein can catalyze integration into DNA sequences present in the human genome that resemble its natural target, the *oriT* (Agundez et al., 2012), opening the possibility that this relaxase could work as a site-specific integrase in human cells (Gonzalez-Prieto et al., 2013). Recently, we have shown that the relaxase TrwC is active in a human cell after delivery by the T4SS of *Bartonella henselae*, where it can promote the integration of foreign DNA into the human genome, although without site-specificity (Gonzalez-Prieto et al., 2017). The integration rate of the foreign DNA introduced by TrwC was about 100 times higher compared to when it was introduced by the Mob relaxase from *Bartonella* cryptic plasmid pRGB1, or by transfection.

Gene therapy strategies combine methods to introduce DNA into specific human cell types and to promote DNA integration in the human genome for stable expression. Bacteria have previously been used as vectors for DNA delivery into mammalian cells; the process, known as bacterofection, is based on the engulfment of bacteria by an eukaryotic cell, which causes bacterial lysis and DNA release (Celec and Gardlik, 2017). We have previously shown that DNA of any origin and length can be introduced into specific human cell types using *B. henselae* as a delivery agent (Fernandez-Gonzalez et al., 2011). In contrast to bacterofection, in this case the DNA is secreted by the living bacterium. *B. henselae* encodes a T4ASS named VirB/D4, which translocates effector proteins to the infected human cell, contributing to its virulence (Saenz et al., 2007). We showed that the VirB/D4 T4SS is also capable of translocating relaxase-DNA complexes via a process resembling bacterial conjugation. DNA transfer was dependent on the conjugative elements required to process the DNA in the donor bacterium, which in this case were derived from the conjugative plasmid R388. No DNA transfer occurred in the absence of the relaxase TrwC, and it was severely impaired in the absence of the conjugative coupling protein TrwB. In a parallel work, Schroder et al. (2011) similarly showed DNA transfer through the *B. henselae* VirB/D4 using the Mob relaxase of a natural plasmid of *Bartonella*; in this case, it was necessary to fuse the known T4 recruiting signal (the BID domain) to the relaxase in order to attain efficient DNA transfer. This discovery had interesting biological implications, opening the possibility that pathogens naturally send DNA to their host cell, and potential biotechnological applications, constituting a new way of DNA delivery to specific human cells (Llosa et al., 2012).

In this work, we asked whether this DNA delivery system could be extended to T4SS from other human pathogens targeting different cell types. We infect cultured mammalian cell lines with *B. henselae*, *L. pneumophila*, or *C. burnetii*, all containing mobilizable plasmids with markers for eukaryotic selection and encoding different conjugative relaxases. We report that DNA can be delivered to human cells through the T4BSS of *L. pneumophila* and *C. burnetii*, which belong to a distant family of T4SS. This suggests that DNA transfer may be an intrinsic feature of T4SS. DNA transfer and integration rates depend on the relaxase used. All these elements could add to the development of useful tools for *in vivo* genetic modification

of human cells. In addition, DNA is a trackable substrate which could be used to study the T4 secretion process in the mammalian host.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Bacterial strains used in this work are listed in **Table 1**. *Escherichia coli* (*Ec*) strains DH5α and D1210 were used for DNA manipulations. *B. henselae* strain RSE247, *L. pneumophila* serogroup 1 strain Lp01 (*hsdR*, *rpsL*; Berger and Isberg, 1993), and *C. burnetii* strain RSA439 Nine Mile phase II (NMII), or derivatives from these strains as indicated, were used for infection of cultured cells.

*Escherichia coli* strains were grown at 37°C in Luria-Bertani broth, supplemented with agar for growth on plates. *B. henselae* was grown on Columbia blood agar (CBA) plates at 37°C under a 5% CO<sub>2</sub> atmosphere. *L. pneumophila* strains were grown on charcoal yeast extract (CYE) plates [1% yeast extract, 1% *N*-(2-acetamido)-2-aminoethanesulfonic acid (ACES; pH 6.9), 3.3 mM L-cysteine, 0.33 mM Fe(NO<sub>3</sub>)<sub>3</sub>, 1.5% Bacto agar, 0.2% activated charcoal] at 37°C, supplemented with 100 μg/ml thymidine if required. *C. burnetii* was grown axenically in liquid acidified citrate cysteine medium 2 (ACCM-2) for 6 days or on ACCM-2 agarose for >8 days at 37°C, 5% CO<sub>2</sub>, and 2.5% O<sub>2</sub> as previously described (Omsland et al., 2011).

For plasmid selection, antibiotics were added at the following final concentrations: ampicillin (Ap), 100 μg/ml; kanamycin monosulfate (Km), 20 μg/ml (*L. pneumophila*), 50 μg/ml (*E. coli*, *B. henselae*) or 375 μg/ml (*C. burnetii*); streptomycin (Sm), 300 μg/ml (*E. coli*) or 100 μg/ml (*B. henselae*, *L. pneumophila*); gentamicin sulfate (Gm), 10 μg/ml (*E. coli*, *B. henselae*) or 5 μg/ml (*L. pneumophila*); chloramphenicol (Cm), 25 μg/ml (*E. coli*) or 3 μg/ml (*C. burnetii*).

Plasmids and Plasmid Constructions

Bacterial plasmids are listed in **Table 2**. Oligonucleotides used for plasmid constructions are listed in **Table 3**. Plasmids pAA58, pLG03, pLG04, pMTX808, pMTX821, and pMTX822 were constructed by the isothermal assembly method (Gibson et al., 2009) using the HiFi assembly cloning kit (New England Biolabs). Plasmids pLG05 and pLG06 were constructed by standard restriction cloning techniques (Sambrook and Russell, 2001).

pAA58 was generated by assembling the eGFP eukaryotic expression cassette from pHP161 into the PstI sites of RSF1010K, which was itself amplified in two overlapping PCR fragments. To generate pLG03, pLG04, pLG05, and pLG06, the hygromycin resistance cassette from pMTX708 was amplified and assembled into the SgsI site of pMTX808 and pAA58, or into the ClaI site of pMTX821 and pMTX822, respectively. pMTX808 was constructed by insertion of an ampicillin resistance cassette (amplified from pJB-KAN) into the *mobA* gene of pAA58. The cassette was inserted at the unique BstZ171 site which lies at nt 320 of *mobA*, leaving unaffected the downstream *mobB* and *repB* ORFs which overlap *mobA*. pMTX821 and pMTX822 were generated by insertion of a kanamycin resistance cassette from pJB-KAN into the gentamicin resistance cassette of pHP159 and pHP181, respectively.

Plasmids were routinely introduced in all strains by electroporation. The protocol for *C. burnetii* electroporation was previously described (Newton et al., 2014); electroporation was carried out with a Bio-Rad GenePulser Xcell (settings: 1.8 kV, 500 Ω, 25 μF). To make competent *L. pneumophila* cells, bacteria were collected from 48 h-patches grown on CYE plates, resuspended in 1 ml ice-cold sterile ddH<sub>2</sub>O, and centrifuged for 2 min in Eppendorf tubes. The washing step was repeated three times. The pellet was resuspended in 1 ml ice-cold sterile glycerol, pelleted for 5 min and resuspended in 1 ml ice-cold sterile glycerol, from which 100 μl aliquots were either frozen at −80°C or used for transformation. Electroporation was

TABLE 1 | Bacterial strains used in this work.

Name	Relevant genotype	Description/comments	Reference
<i>Escherichia coli</i>			
D1210	<i>recA hspR hsdM rpsI lacI<sup>q</sup></i>	Sm <sup>R</sup> , LacI <sup>q</sup> constitutive expression	Sadler et al., 1980
DH5α T1 phage resistant	<i>F- φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk−, mk+) phoA supE44 λ-thi-1 gyrA96 relA1 tonA</i>	Nx <sup>R</sup> , T1 phage resistant strain	Killmann et al., 1996
<i>Bartonella henselae</i>			
RSE247	Sm <sup>R</sup>	Sm <sup>R</sup> spontaneous mutant of ATCC 49882	Schmid et al., 2004
<i>Legionella pneumophila</i>			
Lp02	Lp01 <i>thyA</i>	Spontaneous thymidine auxotroph	Berger and Isberg, 1993
Lp03	Lp02 <i>dotA</i>	Spontaneous <i>dotA</i> mutant	Berger and Isberg, 1993
CR503	Lp01 <i>ΔicmSΔicmW</i>		Coers et al., 2000
<i>Coxiella burnetii</i>			
RSA439	Wild type	Plaque-purified Nine Mile phase II (NMII) clone 4	Williams et al., 1981
RSA439 <i>dotA</i> ::Tn	<i>dotA</i> ::TnA7	Transposon insertion mutant in <i>dotA</i> (CBU_1648), Cm <sup>R</sup> , mCherry	Newton et al., 2014
RSA439 intergenic::Tn	intergenic::TnA7	Transposon insertion mutant between <i>hemD</i> and CBU_2078, Cm <sup>R</sup> , mCherry; shows intracellular replication comparable to wild type	Newton et al., 2014

**TABLE 2 |** Plasmids used in this work.

Relaxase	Other conjugative elements	Plasmid	Selection markers <sup>1</sup>	Description	Reference
Mob-BID	pBGR <i>oriT</i>	pRS130	Km <sup>R</sup> Neo <sup>R</sup>	pBGR:: <i>mob:BID+gfp+neo</i>	(Schroder et al., 2011)
MobA	RSF1010 <i>oriT mobB mobC</i>	RSF1010K	Km <sup>R</sup>	RSF1010 $\Delta$ Sm Km <sup>R</sup>	(Lessl et al., 1993)
MobA	RSF1010 <i>oriT mobB mobC</i>	pAA58	Km <sup>R</sup>	RSF1010K:: <i>egfp</i>	This work
MobA	RSF1010 <i>oriT mobB mobC</i>	pLG04	Km <sup>R</sup> Hyg <sup>R</sup>	pAA58:: <i>hyg</i>	This work
TrwC	R388 <i>oriT trwA trwB</i>	pHP159	Gm <sup>R</sup>	pBBR6:: <i>oriT trwABC+egfp</i>	(Fernandez-Gonzalez et al., 2011)
TrwC	R388 <i>oriT trwA trwB</i>	pHP161	Gm <sup>R</sup>	pBBR6:: <i>oriT trwABC+egfp</i>	(Fernandez-Gonzalez et al., 2011)
TrwC	R388 <i>oriT trwA trwB</i>	pMTX821	Km <sup>R</sup>	pHP159:: <i>Km <math>\Delta</math>Gm</i>	This work
TrwC	R388 <i>oriT trwA trwB</i>	pCOR31	Gm <sup>R</sup> Neo <sup>R</sup>	pHP159:: <i>neo</i>	(Gonzalez-Prieto et al., 2017)
TrwC	R388 <i>oriT trwA trwB</i>	pLG05	Km <sup>R</sup> Hyg <sup>R</sup>	pMTX821:: <i>hyg</i>	This work
TrwC-RalF	R388 <i>oriT trwA trwB</i>	pAA12	Gm <sup>R</sup>	pHP159:: <i>trwC-RalF TS</i>	(Alperi et al., 2013)
–	RSF1010 <i>oriT mobB mobC</i>	pMTX808	Km <sup>R</sup> Ap <sup>R</sup>	pAA58:: <i>Ap MobA-</i>	This work
–	RSF1010 <i>oriT mobB mobC</i>	pLG03	Km <sup>R</sup> Ap <sup>R</sup> Hyg <sup>R</sup>	pMTX808:: <i>hyg</i>	This work
–	R388 <i>oriT trwA trwB</i>	pHP181	Gm <sup>R</sup>	pBBR6:: <i>oriT trwAB+egfp</i>	(Fernandez-Gonzalez et al., 2011)
–	R388 <i>oriT trwA trwB</i>	pMTX822	Km <sup>R</sup>	pHP181:: <i>Km <math>\Delta</math>Gm</i>	This work
–	R388 <i>oriT trwA trwB</i>	pCOR35	Gm <sup>R</sup> Neo <sup>R</sup>	pHP181:: <i>neo</i>	(Gonzalez-Prieto et al., 2017)
–	R388 <i>oriT trwA trwB</i>	pLG06	Km <sup>R</sup> Hyg <sup>R</sup>	pMTX822:: <i>hyg</i>	This work
nr <sup>2</sup>	nr <sup>2</sup>	pMTX708	Ap <sup>R</sup> Hyg <sup>R</sup>	pTRE2hyg:: <i>Ptac-oriT</i>	(Gonzalez-Prieto et al., 2017)
nr <sup>2</sup>	nr <sup>2</sup>	pJB-KAN	Km <sup>R</sup> Ap <sup>R</sup>	Cloning vector	(Omsland et al., 2011)

<sup>1</sup>R, resistance to Ampicillin (Ap), Gentamycin (Gm), Kanamycin (Km), Hygromycin (Hyg) or Neomycin (Neo). <sup>2</sup>nr, not relevant.

**TABLE 3 |** Oligonucleotides used for plasmid constructions.

Plasmid constructed (IA/RC) <sup>1</sup>	Oligonucleotide sequence (5' to 3') <sup>2</sup>	Amplified fragment
pLG03, pLG04 (IA)	TCCAGATGTATGCTCTTCTGCTCGGCGCGCC <b>TTTCGTCTCGAGGCAGTG</b> TGCGATGATAAGCTGTCAAACAGGCGCGCC <b>GTCACTTAGGGTGTGAAAG</b>	Hyg <sup>R</sup> cassette
pLG05, pLG06 (RC)	CCAAACATCGAT <b>GTCAGTTAGGGTGTGAAAG</b> CCAAACATCGAT <b>CTTCGTCTCGAGGCAGTG</b>	Hyg <sup>R</sup> cassette
pAA58 (IA)	<b>AGCTTGCCGCGCCGCGCAG</b> <b>GGTCTATTGCTCCCGGTATTCTGT</b> <b>CGCCAGATCATCGACTTACAGGAATAC</b> <b>GAGCAGAAGAGCATACATCTGGAAGC</b> GCCGCTTTCTGGCTTTGCTTCCAGATGTATGCTCTTCTGCTCGGCGCGCC <b>TGTTTGACAGCTTATCATCGCAG</b> GTGCGGATGAAGTCAGCTCCACCTGCGGCGCGGCAAGCTCCTGCAGG <b>CCCCGACACCCGCCAACAC</b>	RSF1010K    eGFP cassette
pMTX808 (IA)	GCACCTGACCGGTGCCGAGCGCCTGCCGTATT <b>AGAAGGCCATCCTGACGGA</b> TCGCCGCCACCGGCATGGATGGCCAGCGT <b>ATTACCAATGCTTAATCAGTGAG</b>	Ap <sup>R</sup> cassette
pMTX821, pMTX822 (IA)	AGTATGGGCATCATTCGCACATGA <b>AGCGGATTCGCCGCTTTC</b> GGTGGCGGTACTTGGGTGCAT <b>TTATCAGAAGAACTCGTCAAG</b>	Km <sup>R</sup> cassette Km <sup>R</sup> cassette

<sup>1</sup>IA, isothermal assembly; RC, restriction cloning. <sup>2</sup>Nucleotides annealing to the PCR template are shown in bold, and restriction sites used for cloning are underlined.

carried out adding 500 ng DNA and transferring the mixture to a cooled Bio-Rad 0.2-cm cuvette for electroschock with a Bio-Rad GenePulser Xcell set at 2.0 kV, 25  $\mu$ F, and 200  $\Omega$ . After electroporation, 1 ml of AYE broth [1% yeast extract, 1% ACES pH 6.9, 3.3 mM L-cysteine, 0.33 mM Fe(NO<sub>3</sub>)<sub>3</sub>] was added, supplemented with thymidine when required, and the mixture was transferred to a 10 ml tube for incubation for 6 h at 37°C with orbital shaking. The cells were then plated on CYE supplemented with the appropriate antibiotics.

For *B. henselae*, a plate grown for 2 to 3 days was harvested with a sterile cotton swab and resuspended in 950  $\mu$ l of LB. The suspension was centrifuged at 4,000 rpm for 5 min at 4°C, and the pellet was washed in 950  $\mu$ l of ice-cold 10% glycerol (three

times); 40  $\mu$ l of these competent cells was transferred to a cooled tube, and 3  $\mu$ l of DNA (300 ng/ $\mu$ l) was added. The mixture was incubated on ice for 15 min and transferred to a cooled Bio-Rad 0.2-cm cuvette for electroschock with a Bio-Rad Pulse controller II at 2.5 kV/cm, 25  $\mu$ F, and 200  $\Omega$ . After electroporation, 1 ml of SB broth (RPMI 1640 plus L-glutamine, 42 mM HEPES, 1% sodium pyruvate, 5% heat-inactivated fetal calf serum, and 5% sheep blood lysate) was added, and the mixture was transferred to an Eppendorf tube for incubation for 3.5 h at 37°C under 5% CO<sub>2</sub> conditions with slow shaking. The cells were then centrifuged at 4,000 rpm for 4 min at room temperature. The pellet was resuspended in 40  $\mu$ l SB broth and plated on CBA supplemented with the appropriate antibiotics.

### Cell Lines and Cell Culture Conditions

The cell lines used for bacterial infections are listed in **Table 4**. EA.hy926 and HeLa cell lines were routinely grown in Dulbecco’s modified Eagle medium (DMEM; Lonza or Gibco), and Chinese Hamster Ovary (CHO) cells were maintained in minimal essential medium MEM $\alpha$  (Gibco); both media were supplemented with 10% heat inactivated fetal bovine serum (FBS; Lonza or Sigma). Cells were incubated at 37°C under 5% CO<sub>2</sub>.

### Infections

*Bartonella henselae* strains containing the appropriate plasmids were grown on CBA plates for 3 to 4 days. Human cells were seeded 1 day before infection. For routine infections, cells were seeded in 6-well plates (80,000 cells per well) in 3 ml of medium. When the purpose of the infection was to select human cells that had stably acquired the plasmid transferred from *B. henselae*, infections were performed in 10-cm tissue culture dishes seeded with 450,000 cells in 12 ml of medium. The day of infection, DMEM was replaced by M199 medium (Gibco) supplemented with 10% FBS and appropriate antibiotics to select for the *B. henselae* strains to be added. The bacteria were recovered from the CBA plate and resuspended in 1 ml of PBS. The number of bacteria was calculated considering that an OD<sub>600</sub> of 1 corresponds to 10<sup>9</sup> bacteria/ml (Kirby and Nekorchuk, 2002). Bacteria were added to the human cells to get a multiplicity of infection (MOI) of 400 bacteria per host cell. The dishes or plates were incubated for 72 h at 37°C under 5% CO<sub>2</sub>.

*Coxiella burnetii* strains containing the appropriate plasmids were grown for 6 days in liquid cultures. 25,000–50,000 HeLa 229 cells were seeded in DMEM 5% FBS into 24-well plates 6–8 h before they were infected at a MOI of 500, unless specified otherwise. Bacteria were quantified measuring genome equivalents (GE) as previously described (Newton et al., 2014). Infections were incubated for 96 h at 37°C under 5% CO<sub>2</sub>. Wells for quantification of intracellular replication were washed once with PBS at approximately 15 h post infection (hpi) before the addition of fresh DMEM 5% FBS. Wells for flow cytometry experiments were not washed.

*Legionella pneumophila* strains containing the appropriate plasmids were harvested from a heavy patch (after 48 h growth on CYE plates), and used to infect CHO FcγRII cells, stably expressing the receptor FcγRII. This receptor allows

*L. pneumophila* opsonized with anti-*Legionella* antibodies to be internalized efficiently by non-phagocytic cells (Arasaki and Roy, 2010). FcγRII cells were grown to near confluency in 24-well dishes. Bacteria were opsonized with rabbit anti-*Legionella* antibody diluted 1/1000 for 20 min at room temperature with shaking. Bacteria were then added to the cells at an estimated MOI of 10. The cells were centrifuged 5 min at 1000 rpm and incubated for 1 h, washed three times with PBS (Gibco) and incubated in fresh media for 24 h at 37°C under 5% CO<sub>2</sub>.

### Detection of GFP Positive Cells by Flow Cytometry

At the indicated hours post infection (hpi) indicated for each bacteria, infected cells were washed with PBS, trypsinized, and analyzed by flow cytometry using a Cytomics FC500 flow cytometer (Beckman Coulter) for *B. henselae* infections, or a BD Accuri C6 flow cytometer (BD Biosciences) for *L. pneumophila* and *C. burnetii* infections. Data were analyzed using the software for each cytometer and FlowJo (Tree Star, Inc.) software. Singlet cells were gated based on SSC-H/FSC-H and GFP positive cells (detected in the FL1-H channel) were gated based on uninfected control cells. The gate was set to approximately 0.05% GFP<sup>+</sup> cells in the uninfected control sample.

### Fluorescence Microscopy

At the indicated hpi, wells with infected cells were washed with PBS and the plates were placed directly on a Nikon Eclipse TE2000-S inverted fluorescence microscope with a 10× objective lens. Digital images were acquired with a microscope camera (Photometrics CoolSNAP EZ) controlled by SlideBook<sup>TM</sup> (Intelligent Imaging Innovations).

### Detection of Stable Integrants

At 72 hpi, either 500 μg/ml G418 disulfate salt (Sigma–Aldrich) or 300 μg/ml Hygromycin B (Invitrogen), as appropriate, were added to HeLa cells infected with *B. henselae*, and selection was maintained for 4 to 5 weeks. Resistant colonies on the plates were counted.

In order to calculate the integration rate, integration experiments were always performed in parallel with infections to measure GFP positive cells by flow cytometry. The resulting percentage of GFP positive cells was extrapolated to the number of cells in the 10-cm plate used to detect integrants, and the number of resistant colonies was divided by the inferred number of GFP positive cells.

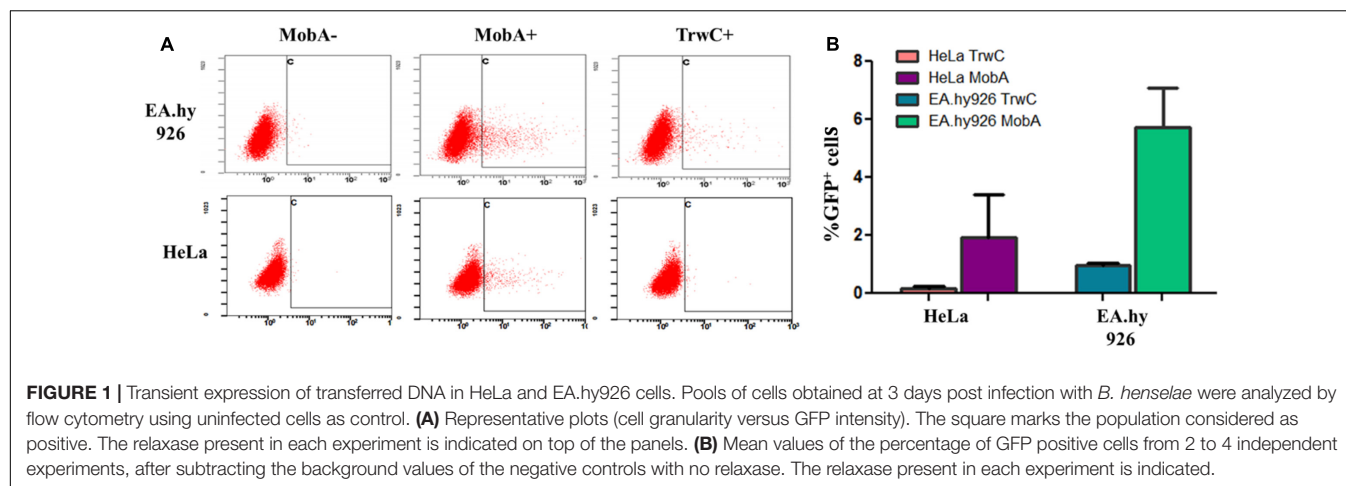
### Determination of Genome Equivalents (GE)

Quantification of *C. burnetii* intracellular replication was performed as described in Newton et al. (2014). Briefly, infected HeLa cells were lysed in ddH<sub>2</sub>O at specific time points post infection. Total genomic DNA was extracted using the Illustra Bacteria GenomicPrep Mini Spin Kit (GE Healthcare) and GE were quantified by qPCR using *dotA*-specific primers (GCGCAATACGCTCAATCACA,

**TABLE 4 |** Mammalian cell lines used in this work.

Name	Description	Reference
CHO FcγRII	Chinese hamster ovary cells producing the FcγRII protein	Joiner et al., 1990
EA.hy926	Fusion cell line of human umbilical vein endothelial cells (HUVEC) and adenocarcinomic human alveolar basal epithelial cells (A549)	ATCC CRL-2922
HeLa	Human epithelial cells of cervix adenocarcinoma	ATCC CCL-2
HeLa 229	Human epithelial cells of cervix adenocarcinoma	ATCC CCL-2.1





CCATGGCCCCAATTCTCTT). The generation of this short PCR product is not affected by the presence of a transposon in the *dotA::Tn* mutant strain.

## RESULTS

The conjugative relaxase TrwC can be translocated through the T4SS VirB/D4 of *B. henselae* to human cells, where it promotes the integration of the transferred DNA into the recipient genome (Gonzalez-Prieto et al., 2017). In this work, we wanted to test whether this is a unique feature of TrwC and VirB/D4, or other systems can also be combined to deliver and integrate DNA into human cells.

To test DNA transfer mediated by the relaxase MobA of the mobilizable plasmid RSF1010, we constructed a derivative carrying an eukaryotic eGFP expression cassette to detect gene expression from the human cell nucleus. An insertion of an ampicillin resistance cassette in *mobA* served as a negative control. The insertion is located in the 5' region of the ORF, thus not affecting the expression of the ORFs *mobB* and especially *repB*, which encodes a DNA primase required for plasmid replication. We observed that this *mobA*<sup>-</sup> construct had a higher copy number than the parental plasmid, as judged from the amount of DNA extracted from parallel cultures (data not shown). This phenomenon has previously been reported, and attributed to the repressor role of MobA/RepB in replication (Frey et al., 1992).

These plasmids (pAA58 and pMTX808; Table 2) were introduced in *B. henselae*, and the resulting strains were used to infect both EA.hy926 and HeLa human cell lines. The former is derived from HUVEC cells, which are the natural target of *B. henselae* *in vivo*; however, HeLa cells can also be infected by *B. henselae* with lower efficiency, and we showed that TrwC-mediated DNA transfer takes place to HeLa cells as well (Gonzalez-Prieto et al., 2017). *B. henselae* carrying plasmids coding for either MobA or TrwC, or relaxase mutants as negative controls, were used for infections. To assess transfer of the plasmid DNA to the human cells, flow cytometry was used to quantify the expression of the eGFP cassette per

cell, thus allowing the determination of the percentage of GFP positive cells. The results are shown in Figure 1 and Table 5, top 8 rows. We observed DNA transfer when the plasmids encoded a functional relaxase, and background levels in the absence of a relaxase. DNA transfer rates were notably higher when using MobA as the leading relaxase compared to TrwC.

In order to measure genomic integration of the transferred DNA, we constructed plasmid derivatives encoding antibiotic resistance cassettes (see Table 2). The plasmids containing R388 conjugative elements carried a neomycin gene; however, this was not used in these experiment because of the presence of a kanamycin resistance gene in the RSF1010K backbone, which could lead to recombination between both cassettes. Instead, a hygromycin resistance cassette was inserted. In order to avoid an effect caused by the different antibiotic selections applied, we also constructed Hygromycin-resistant derivatives encoding TrwC (Table 2), and we found that TrwC-mediated integration rate did not vary when the selection applied was hygromycin B or Geneticin (data not shown).

HeLa cells were used as target cells to measure DNA integration, because in contrast to EA.hy926 cells HeLa cells show enhanced survival during the 4–5 weeks of antibiotic selection required to measure resistant colonies (Gonzalez-Prieto et al., 2017). The cells were infected with *B. henselae* carrying the different plasmids. A plasmid derived from the cryptic *Bartonella* plasmid pBGR1 was also assayed for comparison, since it has been reported that its relaxase mediates DNA transfer but does not promote integration of the transferred DNA (Gonzalez-Prieto et al., 2017). After applying the antibiotic selection, resistant colonies were counted, and integration rates were calculated dividing this number by the number of GFP positive cells determined in parallel infection experiments (see Materials and Methods for details). The results (Figure 2) indicate that the integration rate for the MobA constructs was approximately one-log higher than in case of Mob-BID, which suggest that MobA promotes integration of the transferred DNA. It can also be observed that TrwC has a stronger effect on integration than MobA (approximately five-fold higher DNA integration).

**TABLE 5 |** Rates of DNA transfer to mammalian cells through T4ASS and T4BSS.

Donor bacteria (genotype)	T4SS	Transfer system	Relaxase	Infected cells	GFP <sup>+</sup> mammalian cells <sup>(1)</sup>	
					Flow cyt %	Scope
<i>Bh</i> RSE247 (wt)	Functional	RSF1010	MobA	EA.hy926	5.72 ± 1.37	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	RSF1010	–	EA.hy926	0.29 ± 0.07	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	R388	TrwC	EA.hy926	1.00 ± 0.09	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	R388	–	EA.hy926	0.14 ± 0.19	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	RSF1010	MobA	HeLa	2.00 ± 1.48	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	RSF1010	–	HeLa	0.07 ± 0.05	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	R388	TrwC	HeLa	0.20 ± 0.03	nq <sup>(2)</sup>
<i>Bh</i> RSE247 (wt)	Functional	R388	–	HeLa	0.04 ± 0.06	nq <sup>(2)</sup>
<i>Lp</i> Lp02 (wt)	Functional	RSF1010	MobA	CHO FcyRII	0.35 ± 0.12	nq <sup>(2)</sup>
<i>Lp</i> Lp03 ( <i>dotA</i> )	No transport	RSF1010	MobA	CHO FcyRII	0.03 ± 0.05	<5 × 10 <sup>−6</sup>
<i>Lp</i> Lp02 (wt)	Functional	RSF1010	–	CHO FcyRII	0.00 ± 0.00	<5 × 10 <sup>−6</sup>
<i>Lp</i> CR503 ( <i>icmS icmW</i> )	No chaperone	RSF1010	MobA	CHO FcyRII	0.00 ± 0.00	<5 × 10 <sup>−6</sup>
<i>Lp</i> Lp02 (wt)	Functional	R388	TrwC	CHO FcyRII	0.00 ± 0.00	<5 × 10 <sup>−6</sup>
<i>Lp</i> Lp02 (wt)	Functional	R388	TrwC-RalF	CHO FcyRII	0.00 ± 0.00	1 × 10 <sup>−5</sup>
<i>Lp</i> Lp03 ( <i>dotA</i> )	No transport	R388	TrwC-RalF	CHO FcyRII	nq <sup>(2)</sup>	<5 × 10 <sup>−6</sup>
<i>Lp</i> Lp02 (wt)	Functional	R388	–	CHO FcyRII	nq <sup>(2)</sup>	<5 × 10 <sup>−6</sup>
<i>Lp</i> CR503 ( <i>icmS icmW</i> )	No chaperone	R388	TrwC-RalF	CHO FcyRII	nq <sup>(2)</sup>	2 × 10 <sup>−5</sup>
<i>Cb</i> intergenic::Tn (wt)	Functional	RSF1010	MobA	HeLa	0.56 ± 0.53	nq <sup>(2)</sup>
<i>Cb</i> <i>dotA</i> ::Tn	No transport	RSF1010	MobA	HeLa	0.04 ± 0.02	nq <sup>(2)</sup>
<i>Cb</i> intergenic::Tn (wt)	Functional	RSF1010	–	HeLa	0.10 <sup>(3)</sup> ± 0.04	<5 × 10 <sup>−6</sup>

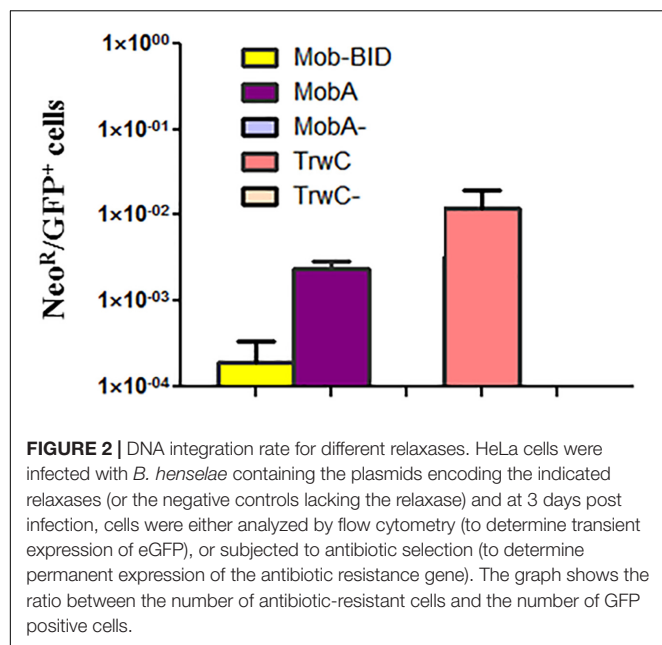
<sup>(1)</sup>DNA transfer is measured as the ratio of mammalian recipient cells expressing GFP. Data from flow cytometry (left column) show the percentage of GFP positive cells (mean ± SD of two to eight independent assays). Infected cells were also screened visually under the microscope (right column). Positive cells were counted and divided by the total number of cells per well (estimated as 200,000). The screen was performed at least twice for each condition. <sup>(2)</sup>nq, not quantified. <sup>(3)</sup>Due to higher background (see text for details).

Earlier studies reported Dot/Icm-dependent conjugative DNA transfer of RSF1010 (Vogel et al., 1998), implying that MobA can mediate the translocation of an attached DNA substrate through the T4BSS Dot/Icm of *L. pneumophila*. Thus, we asked whether the Dot/Icm T4SS could also promote DNA transfer to mammalian cells upon infection by *L. pneumophila*. In addition to testing MobA-mediated transfer, we tested DNA transfer mediated by TrwC and TrwC-RalF, a fusion protein carrying the C-terminal 20 residues of the *L. pneumophila* Dot/Icm substrate RalF, that has been shown to be sufficient for translocation (Nagai et al., 2005). In contrast to the infection experiments done with *B. henselae*, for infections with *L. pneumophila* a MOI of 10 was used and DNA transfer was monitored at 24 hpi. As shown in **Figure 3A** and **Table 5**, we detected GFP positive cells after infection by a mechanism dependent on the Dot/Icm T4BSS and the relaxase MobA. Thus, we show for the first time that DNA transfer can occur through a T4BSS into mammalian cells. Using the same flow cytometry assay, we did not detect GFP positive cells above the background when the mobilizable plasmids encoded the relaxase TrwC or TrwC-RalF. However, inspection of the infected cells by fluorescence microscopy did reveal a small number of positive cells that expressed GFP uniformly and strongly after infection with *L. pneumophila* producing TrwC-RalF (**Figure 3B**). Positive cells were not observed in the negative controls or with TrwC-encoding plasmids.

The rate of DNA transfer was highly dependent on the conjugative DNA processing system used. This could be due to different relaxase recruitment efficiencies. The Dot/Icm T4BSS recruits a subset of its substrates through a chaperone complex formed by IcmS and IcmW (Cambronne and Roy, 2007). To determine if recruitment of the relaxases was dependent on this complex, a *L. pneumophila*  $\Delta icmS \Delta icmW$  mutant strain was used in infection experiments carrying plasmids which encode either MobA or TrwC-RalF. The results (**Table 5** and **Figure 3B**) indicate that the absence of IcmSW did not affect DNA transfer mediated by TrwC-RalF, while DNA transfer mediated by MobA was abolished in the absence of IcmSW.

The Dot/Icm T4BSS of *L. pneumophila* is closely related to that of *C. burnetii*, and several reports have shown that both can recruit the same effector proteins and cross-complement *icmSW* mutants (Zamboni et al., 2003; Zusman et al., 2003; Carey et al., 2011). Thus, we decided to test MobA-mediated DNA transfer through the Dot/Icm T4BSS of *C. burnetii*. HeLa cells were infected with *C. burnetii* strains harboring the plasmids with and without MobA at a MOI of 500, and GFP expression was investigated at 4 days post infection. The results are shown in **Table 5**, and **Figure 3C** shows representative plots. Similar to what was observed with *L. pneumophila*, GFP positive cells were only detected when the Dot/Icm T4BSS and the MobA relaxase were present.

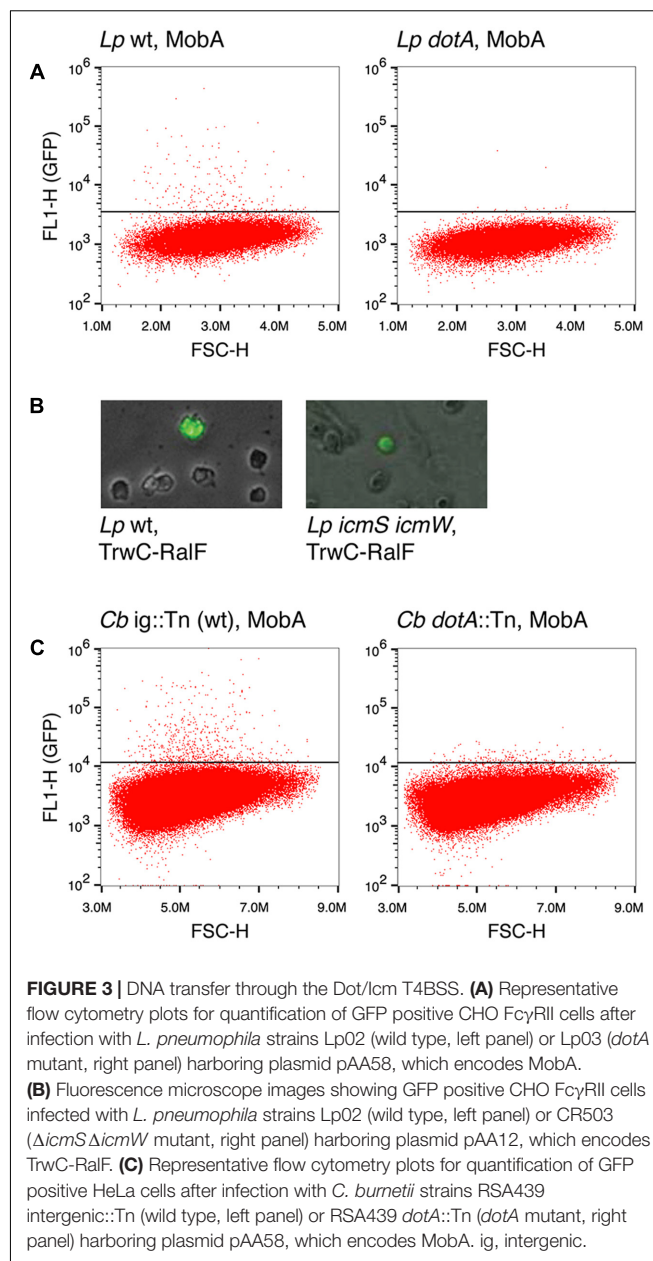
Performing these experiments, we observed a difference in the background fluorescence intensity of HeLa cells depending



on the bacterial strain used for infection. A representative flow cytometry histogram is shown in **Figure 4A**. The background GFP fluorescence peak shifts toward a higher intensity when HeLa cells were infected with wild type *C. burnetii* or wild type *C. burnetii* harboring the plasmid with the *mobA* mutation, but not when cells were infected with wild type *C. burnetii* carrying the plasmid with the intact *mobA* gene. This higher fluorescence did not correspond to DNA transfer, since we did not detect any proper GFP positive cells by flow cytometry or using microscopy, but it contributed to a minimal raise in the background frequencies observed when infecting with a *mobA*<sup>-</sup> strain (see **Table 5**). However, the difference in background fluorescence may be attributed to a different amount of intracellular bacteria per cell. To test this hypothesis, HeLa cells were infected at a MOI of 50 and the number of intracellular *C. burnetii* was determined by measuring GE at two time points post infection. The results are shown in **Figure 4B**. A strain carrying the *mobA*-deficient plasmid replicates nearly as efficiently as a strain with no plasmid. In contrast, the same strain carrying a plasmid that encodes a functional MobA protein was severely impaired in intracellular replication. A *dotA* mutant that fails to replicate intracellularly due to the absence of a functional T4SS was used as a control in this assay.

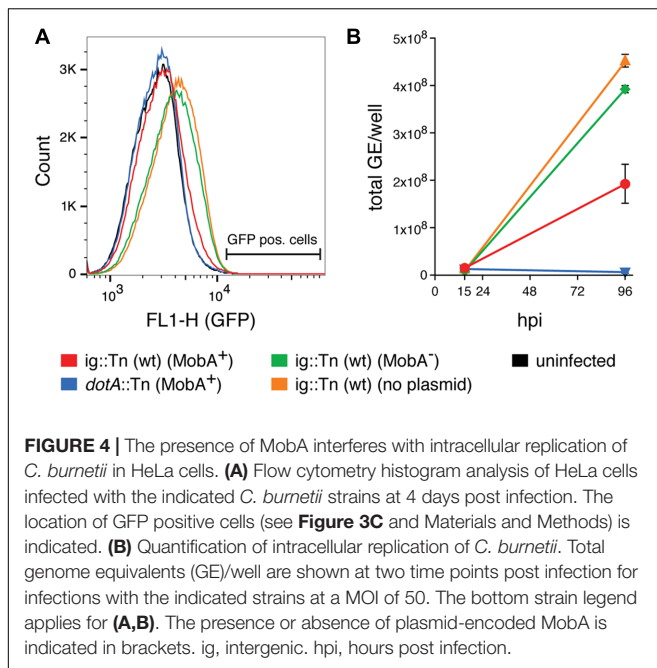
## DISCUSSION

In our previous reports, we showed that the conjugative relaxase TrwC can be translocated to human cells through the T4SS VirB/D4 of *B. henselae* (Fernandez-Gonzalez et al., 2011), and also that it promotes integration of the transferred DNA into the recipient genome (Gonzalez-Prieto et al., 2017). Whether these abilities were unique for TrwC and VirB/D4 remained to be tested. In this work, we report that different relaxases



and T4SS can be used to transfer DNA to human cells and to promote DNA integration. In other words, relaxases and T4SS from various bacterial species can be combined to create tools intended to genetically modify specific human target cells in a permanent way, thus generating enormous biotechnological potential.

Firstly, we compared the ability of different relaxases to transfer DNA to mammalian cells and to promote DNA integration into the recipient genome when translocated by the same T4SS, VirB/D4. Human cells were infected with *B. henselae* carrying derivatives of the mobilizable plasmid RSF1010, encoding the relaxase MobA; with constructs containing the conjugative processing elements of the self-transferable plasmid R388, which encodes the relaxase TrwC; or with derivatives



of *B. henselae* cryptic plasmid pBGR1, coding for the relaxase Mob fused to the BID signal for efficient recruitment by VirB/D4 (Schroder et al., 2011). When the three plasmids are compared in terms of DNA transfer and integration rates (Figures 1, 2), we find that these vary significantly, with RSF1010 being the most efficiently transferred, while TrwC is the relaxase showing higher integration rates. The rate of DNA transfer is probably proportional to the efficiency with which the relaxase is recruited to the T4SS machinery; this assumption comes from previous works showing that the relaxase Mob itself could transfer DNA to human cells with barely detectable frequency, but when a recruitment secretion signal was fused to its C-terminal end, it transferred DNA to similar frequencies than TrwC (Schroder et al., 2011). In addition, in case of R388, a deletion of the conjugative coupling protein, a component believed to play a key role in the recruitment of the conjugative substrate, caused DNA transfer rates to drop 10-fold (Fernandez-Gonzalez et al., 2011). The relaxase MobA belongs to a mobilizable plasmid which hijacks the T4SS of co-residing conjugative plasmids, so it can be translocated through various T4SS; thus, it is plausible that the requirements for MobA recruitment are less stringent. In fact, the C-terminal 48 residues of MobA were shown to direct translocation of a Cre fusion through the VirB T4SS of *A. tumefaciens* into plant cells (Vergunst et al., 2005). Now, we show that MobA can also be translocated through a T4ASS into mammalian cells.

The ability to enhance integration of the transferred DNA into the recipient cell genome must reside in an intrinsic property of the relaxase, which is the only protein entering the recipient cell covalently attached to the transferred DNA strand. We report here that the promiscuous relaxase MobA

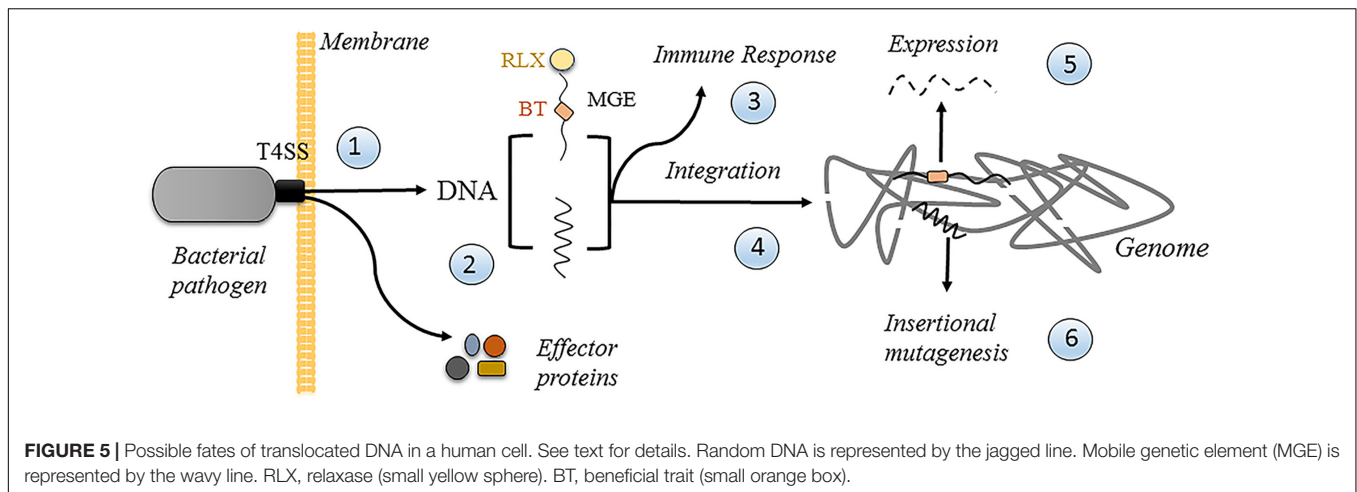
also promotes DNA integration, resulting in resistant colonies with about 10-fold higher frequency than Mob-BID, which does not promote integration above background levels obtained by DNA transfection (Gonzalez-Prieto et al., 2017), but roughly five-fold lower frequency than TrwC. These differences observed among relaxases could be due to differential nuclear targeting, catalytic activity, or binding affinity to its target, which could protect the DNA ends, thus favoring integration by host-mediated mechanisms, as previously suggested (Gonzalez-Prieto et al., 2017). Subcellular localization of TrwC and MobA in human cells showed no preferential nuclear localization for either relaxase (Silby et al., 2007; Agundez et al., 2011). It is noteworthy that TrwC catalyzes site-specific recombination on supercoiled DNA substrates (Cesar et al., 2006), while MobA was shown to catalyze site-specific recombination between two *oriT* copies when the substrate was single-stranded (Meyer, 1989), and other relaxases do not catalyze this reaction at all. Although the integration pattern in the human genome is random (Gonzalez-Prieto et al., 2017), site-specific recombination ability could play a role in strand-transfer reactions when the nucleoprotein complex is directed to a nicked DNA strand by the host repair machinery.

MobA can be translocated by the T4BSS of *L. pneumophila*, alone or bound to DNA, into recipient bacteria (Vogel et al., 1998; Luo and Isberg, 2004). These results prompted us to test its translocation by T4BSS into mammalian cells. Our results (Figure 3) show for the first time that DNA transfer to human cells can also be accomplished through the Dot/Icm T4BSS of *L. pneumophila* and *C. burnetii*, only remotely related to T4ASS. Thus, it is reasonable to assume that DNA translocation may be an intrinsic ability of T4SS. An important difference between both Dot/Icm systems is the temporal pattern of secretion: while *L. pneumophila* has been shown to secrete effectors as internalization into host cells is initiated (Nagai et al., 2005) in case of *C. burnetii* effector translocation is initiated when the pathogen has reached an acidified lysosomal compartment (Newton et al., 2013); thus, DNA transfer in *C. burnetii* must occur from within the *Coxiella*-containing vacuole.

DNA transfer was dependent on the presence of the Dot/Icm T4SS and a functional relaxase, as expected for a *bona fide* conjugation-like DNA transfer process. The wide differences in DNA transfer rates depending on the relaxase (MobA, TrwC, or TrwC-RalF, including the translocation signal of the natural T4SS substrate RalF) and on the presence/absence of the chaperones IcmSW (see Table 5) support the concept that relaxase recruitment is the main driver of DNA transfer.

During the course of performing *C. burnetii* infection experiments, we noticed an inhibition of *C. burnetii* intracellular replication caused by the presence of RSF1010 derivatives carrying a functional MobA relaxase while isogenic strains with a *mobA* mutation did not affect growth (Figure 4). Similarly, RSF1010 conjugation was shown to inhibit intracellular replication and virulence of *L. pneumophila* (Segal and Shuman, 1998), probably by MobA interference with effector secretion by Dot/Icm. This result should be taken into account when using vectors based on RSF1010, which are the more commonly used by both *L. pneumophila* and *C. burnetii*.





Finally, an attractive question that remains open is the possible biological role, if any, of DNA transfer to mammalian cells by bacterial pathogens harboring a T4SS. Is the DNA transfer ability an evolutionary remnant of the conjugative T4SS from which the T4SS involved in virulence probably have evolved? Or is it an ability which the pathogens have evolved to use to their own benefit, in the same way as *A. tumefaciens* uses it to subvert its eukaryotic host cell?

In support of the first possibility, it is relevant to point out that in spite of many attempts, no T4 protein, protein domain or amino acid residue has been identified to date, which is specifically involved in DNA transfer. All analyzed mutants in T4 components, even in the conjugative coupling protein ATPase, affected DNA and protein translocation to the same extent, leading to the suggestion that relaxase and DNA translocation may have the same molecular requirements (de Paz et al., 2010; Larrea et al., 2017). Thus, the ability to transfer DNA could not be lost in a T4SS even if it evolved to only secrete proteins. However, the potential of DNA transfer for long-term subversion of the host cells makes it attractive to think that pathogens may utilize such a process for their own profit. **Figure 5** illustrates the possible fates of secreted DNA in a human cell. A pathogen translocates effector proteins and DNA through its T4SS once in contact with the membrane (1 in **Figure 5**), whether it is from within a vacuolar compartment, as in case of *C. burnetii*, or from the outside. The secreted DNA could either be random DNA, as proposed for *H. pylori* (Varga et al., 2016), or a specifically recruited mobile genetic element (MGE), in which case a dedicated transfer system would attach a relaxase to its end (2). The cytoplasmic DNA could elicit an immune response (3), as proposed for *H. pylori* (Varga et al., 2016), which could be used by the pathogen for its own benefit. DNA could also get integrated into the host cell genome (4) by the host repair/recombination systems, and/or by the covalently attached conjugative relaxase. Integration will lead to the stable expression of the encoded information (5), including any beneficial traits that the pathogen may have evolved to encode in MGE for that purpose. Finally, random integration has an inherent risk of insertional mutagenesis (6), which could lead to increased growth

of the host cell, thereby promoting the extension of the niche of the pathogen.

In this context, it has to be stressed that human pathogens contain many poorly characterized MGE, which could be substrates for DNA transfer (in addition to the possibility of sporadic transfer of visiting promiscuous plasmids, such as RSF1010). As examples from the pathogens used in this study, the pBRG1 cryptic plasmid of *B. henselae* can be recruited by VirB/D4 and translocated to human cells (Schroder et al., 2011); conjugative transfer of chromosomal DNA has been reported for *L. pneumophila* (Miyamoto et al., 2003), and its genome includes several genomic islands; and notably, a cryptic plasmid in *C. burnetii* is enriched in important effector genes (Voth et al., 2011); it is tempting to speculate that this plasmid may be transferred to the host cell.

## AUTHOR CONTRIBUTIONS

All authors contributed to the conception and design of the work, data acquisition and/or analysis. All authors contributed to drafting, revising, and final approval of the work. All authors agree to be accountable for all aspects of the work.

## FUNDING

This work was supported by grant BIO2013-46414-P from the Spanish Ministry of Economy and Competitiveness to ML, and NIH grants AI041699 and AI114760 to CRR. DLG was supported by a predoctoral fellowship from the University of Cantabria (Spain). SS was supported by an Advanced Postdoc Mobility fellowship from the Swiss National Science Foundation (SNSF).

## ACKNOWLEDGMENT

ML wishes to thank the Roy lab, and especially David Chetrit and Stephanie Shames, for their support with *Legionella* lab protocols.

## REFERENCES

- Agundez, L., Gonzalez-Prieto, C., Machon, C., and Llosa, M. (2012). Site-specific integration of foreign DNA into minimal bacterial and human target sequences mediated by a conjugative relaxase. *PLoS ONE* 7:e31047. doi: 10.1371/journal.pone.0031047
- Agundez, L., Machon, C., Cesar, C. E., Rosa-Garrido, M., Delgado, M. D., and Llosa, M. (2011). Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. *Appl. Environ. Microbiol.* 77, 201–210. doi: 10.1128/AEM.01371-10
- Alperi, A., Larrea, D., Fernandez-Gonzalez, E., Dehio, C., Zechner, E. L., and Llosa, M. (2013). A translocation motif in relaxase TrwC specifically affects recruitment by its conjugative type IV secretion system. *J. Bacteriol.* 195, 4999–5006. doi: 10.1128/JB.00367-13
- Arasaki, K., and Roy, C. R. (2010). *Legionella pneumophila* promotes functional interactions between plasma membrane syntaxins and Sec22b. *Traffic* 11, 587–600. doi: 10.1111/j.1600-0854.2010.01050.x
- Berger, K. H., and Isberg, R. R. (1993). Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila*. *Mol. Microbiol.* 7, 7–19. doi: 10.1111/j.1365-2958.1993.tb01092.x
- Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* 39, 81–95. doi: 10.1111/1574-6976.12085
- Cambronne, E. D., and Roy, C. R. (2007). The *Legionella pneumophila* IcmSW complex interacts with multiple Dot/Icm effectors to facilitate type IV translocation. *PLoS Pathog.* 3:e188. doi: 10.1371/journal.ppat.0030188
- Carey, K. L., Newton, H. J., Luhrmann, A., and Roy, C. R. (2011). The *Coxiella burnetii* Dot/Icm system delivers a unique repertoire of type IV effectors into host cells and is required for intracellular replication. *PLoS Pathog.* 7:e1002056. doi: 10.1371/journal.ppat.1002056
- Celec, P., and Gardlik, R. (2017). Gene therapy using bacterial vectors. *Front. Biosci.* 22:81–95. doi: 10.2741/4473
- Cesar, C. E., Machon, C., de la Cruz, F., and Llosa, M. (2006). A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Mol. Microbiol.* 62, 984–996. doi: 10.1111/j.1365-2958.2006.05437.x
- Chandran Darbari, V., and Waksman, G. (2015). Structural biology of bacterial type IV secretion systems. *Annu. Rev. Biochem.* 84, 603–629. doi: 10.1146/annurev-biochem-062911-102821
- Christie, P. J. (2016). The mosaic type IV secretion systems. *EcoSal Plus* 7, doi: 10.1128/ecosalplus.ESP-0020-2015
- Coers, J., Kagan, J. C., Matthews, M., Nagai, H., Zuckman, D. M., and Roy, C. R. (2000). Identification of Icm protein complexes that play distinct roles in the biogenesis of an organelle permissive for *Legionella pneumophila* intracellular growth. *Mol. Microbiol.* 38, 719–736. doi: 10.1046/j.1365-2958.2000.02176.x
- de Paz, H. D., Larrea, D., Zunzunegui, S., Dehio, C., de la Cruz, F., and Llosa, M. (2010). Functional dissection of the conjugative coupling protein TrwB. *J. Bacteriol.* 192, 2655–2669. doi: 10.1128/JB.01692-09
- Draper, O., Cesar, C. E., Machon, C., de la Cruz, F., and Llosa, M. (2005). Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16385–16390. doi: 10.1073/pnas.0506081102
- Fernandez-Gonzalez, E., de Paz, H. D., Alperi, A., Agundez, L., Faustmann, M., Sangari, F. J., et al. (2011). Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *J. Bacteriol.* 193, 6257–6265. doi: 10.1128/JB.05905-11
- Frey, J., Bagdasarian, M. M., and Bagdasarian, M. (1992). Replication and copy number control of the broad-host-range plasmid RSF1010. *Gene* 113, 101–106. doi: 10.1016/0378-1119(92)90675-F
- Garcillan-Barcia, M. P., Jurado, P., Gonzalez-Perez, B., Moncalian, G., Fernandez, L. A., and de la Cruz, F. (2007). Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. *Mol. Microbiol.* 63, 404–416. doi: 10.1111/j.1365-2958.2006.05523.x
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Gonzalez-Perez, B., Lucas, M., Cooke, L. A., Vyle, J. S., de la Cruz, F., and Moncalian, G. (2007). Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. *EMBO J.* 26, 3847–3857. doi: 10.1038/sj.emboj.7601806
- Gonzalez-Prieto, C., Agundez, L., Linden, R. M., and Llosa, M. (2013). HUH site-specific recombinases for targeted modification of the human genome. *Trends Biotechnol.* 31, 305–312. doi: 10.1016/j.tibtech.2013.02.002
- Gonzalez-Prieto, C., Gabriel, R., Dehio, C., Schmidt, M., and Llosa, M. (2017). The conjugative relaxase TrwC promotes integration of foreign DNA in the human genome. *Appl. Environ. Microbiol.* 83:e00207-17. doi: 10.1128/AEM.00207-17
- Hubber, A., and Roy, C. R. (2010). Modulation of host cell function by *Legionella pneumophila* type IV effectors. *Annu. Rev. Cell Dev. Biol.* 26, 261–283. doi: 10.1146/annurev-cellbio-100109-104034
- Joiner, K. A., Fuhrman, S. A., Miettinen, H. M., Kasper, L. H., and Mellman, I. (1990). *Toxoplasma gondii*: fusion competence of parasitophorous vacuoles in Fc receptor-transfected fibroblasts. *Science* 249, 641–646. doi: 10.1126/science.2200126
- Killmann, H., Benz, R., and Braun, V. (1996). Properties of the FhuA channel in the *Escherichia coli* outer membrane after deletion of FhuA portions within and outside the predicted gating loop. *J. Bacteriol.* 178, 6913–6920. doi: 10.1128/jb.178.23.6913-6920.1996
- Kirby, J. E., and Nekorchuk, D. M. (2002). *Bartonella*-associated endothelial proliferation depends on inhibition of apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4656–4661. doi: 10.1073/pnas.072292699
- Larrea, D., de Paz, H. D., Matilla, I., Guzman-Herrador, D. L., Lasso, G., de la Cruz, F., et al. (2017). Substrate translocation involves specific lysine residues of the central channel of the conjugative coupling protein TrwB. *Mol. Genet. Genomics* doi: 10.1007/s00438-017-1331-3 [Epub ahead of print].
- Lessl, M., Balzer, D., Weyrauch, K., and Lanka, E. (1993). The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. *J. Bacteriol.* 175, 6415–6425. doi: 10.1128/jb.175.20.6415-6425.1993
- Llosa, M., Bolland, S., Grandoso, G., and de la Cruz, F. (1994). Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC. *J. Bacteriol.* 176, 3210–3217. doi: 10.1128/jb.176.11.3210-3217.1994
- Llosa, M., Schroder, G., and Dehio, C. (2012). New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol.* 20, 355–359. doi: 10.1016/j.tim.2012.05.008
- Luo, Z. Q., and Isberg, R. R. (2004). Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc. Natl. Acad. Sci. U.S.A.* 101, 841–846. doi: 10.1073/pnas.0304916101
- Meyer, R. (1989). Site-specific recombination at oriT of plasmid R1162 in the absence of conjugative transfer. *J. Bacteriol.* 171, 799–806. doi: 10.1128/jb.171.2.799-806.1989
- Miyamoto, H., Yoshida, S., Taniguchi, H., and Shuman, H. A. (2003). Virulence conversion of *Legionella pneumophila* by conjugal transfer of chromosomal DNA. *J. Bacteriol.* 185, 6712–6718. doi: 10.1128/JB.185.22.6712-6718.2003
- Nagai, H., Cambronne, E. D., Kagan, J. C., Amor, J. C., Kahn, R. A., and Roy, C. R. (2005). A C-terminal translocation signal required for Dot/Icm-dependent delivery of the *Legionella* RalF protein to host cells. *Proc. Natl. Acad. Sci. U.S.A.* 102, 826–831. doi: 10.1073/pnas.0406239101
- Newton, H. J., Kohler, L. J., McDonough, J. A., Temoche-Diaz, M., Crabill, E., Hartland, E. L., et al. (2014). A screen of *Coxiella burnetii* mutants reveals important roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. *PLoS Pathog.* 10:e1004286. doi: 10.1371/journal.ppat.1004286
- Newton, H. J., McDonough, J. A., and Roy, C. R. (2013). Effector protein translocation by the *Coxiella burnetii* Dot/Icm type IV secretion system requires endocytic maturation of the pathogen-occupied vacuole. *PLoS ONE* 8:e54566. doi: 10.1371/journal.pone.0054566
- Omstand, A., Beare, P. A., Hill, J., Cockrell, D. C., Howe, D., Hansen, B., et al. (2011). Isolation from animal tissue and genetic transformation of *Coxiella*

- burnetii* are facilitated by an improved axenic growth medium. *Appl. Environ. Microbiol.* 77, 3720–3725. doi: 10.1128/AEM.02826-10
- Perssonic, N., Barlocher, K., Finsel, I., and Hilbi, H. (2016). Subversion of retrograde trafficking by translocated pathogen effectors. *Trends Microbiol.* 24, 450–462. doi: 10.1016/j.tim.2016.02.003
- Rolando, M., and Buchrieser, C. (2014). *Legionella pneumophila* type IV effectors hijack the transcription and translation machinery of the host cell. *Trends Cell Biol.* 24, 771–778. doi: 10.1016/j.tcb.2014.06.002
- Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. *Gene* 8, 279–300. doi: 10.1016/0378-1119(80)90005-0
- Saenz, H. L., Engel, P., Stoeckli, M. C., Lanz, C., Raddatz, G., Vayssier-Taussat, M., et al. (2007). Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. *Nat. Genet.* 39, 1469–1476. doi: 10.1038/ng.2007.38
- Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Schmid, M. C., Schulein, R., Dehio, M., Denecker, G., Carena, I., and Dehio, C. (2004). The VirB type IV secretion system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. *Mol. Microbiol.* 52, 81–92. doi: 10.1111/j.1365-2958.2003.03964.x
- Schroder, G., Schulein, R., Quebatte, M., and Dehio, C. (2011). Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen *Bartonella henselae*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14643–14648. doi: 10.1073/pnas.1019074108
- Segal, G., and Shuman, H. A. (1998). Intracellular multiplication and human macrophage killing by *Legionella pneumophila* are inhibited by conjugal components of IncQ plasmid RSF1010. *Mol. Microbiol.* 30, 197–208. doi: 10.1046/j.1365-2958.1998.01054.x
- Silby, M. W., Ferguson, G. C., Billington, C., and Heinemann, J. A. (2007). Localization of the plasmid-encoded proteins TraI and MobA in eukaryotic cells. *Plasmid* 57, 118–130. doi: 10.1016/j.plasmid.2006.08.006
- Varga, M. G., Shaffer, C. L., Sierra, J. C., Suarez, G., Piazuelo, M. B., Whitaker, M. E., et al. (2016). Pathogenic *Helicobacter pylori* strains translocate DNA and activate TLR9 via the cancer-associated cag type IV secretion system. *Oncogene* 35, 6262–6269. doi: 10.1038/ncr.2016.158
- Vergunst, A. C., van Lier, M. C., den Dulk-Ras, A., Stuve, T. A., Ouwehand, A., and Hooykaas, P. J. (2005). Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 832–837. doi: 10.1073/pnas.0406241102
- Vogel, J. P., Andrews, H. L., Wong, S. K., and Isberg, R. R. (1998). Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279, 873–876. doi: 10.1126/science.279.5352.873
- Voth, D. E., Beare, P. A., Howe, D., Sharma, U. M., Samoilis, G., Cockrell, D. C., et al. (2011). The *Coxiella burnetii* cryptic plasmid is enriched in genes encoding type IV secretion system substrates. *J. Bacteriol.* 193, 1493–1503. doi: 10.1128/JB.01359-10
- Williams, J. C., Peacock, M. G., and McCaul, T. F. (1981). Immunological and biological characterization of *Coxiella burnetii*, phases I and II, separated from host components. *Infect. Immun.* 32, 840–851.
- Zamboni, D. S., McGrath, S., Rabinovitch, M., and Roy, C. R. (2003). *Coxiella burnetii* express type IV secretion system proteins that function similarly to components of the *Legionella pneumophila* Dot/Icm system. *Mol. Microbiol.* 49, 965–976. doi: 10.1046/j.1365-2958.2003.03626.x
- Zechner, E. L., Lang, S., and Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1073–1087. doi: 10.1098/rstb.2011.0207
- Zusman, T., Yerushalmi, G., and Segal, G. (2003). Functional similarities between the icm/dot pathogenesis systems of *Coxiella burnetii* and *Legionella pneumophila*. *Infect. Immun.* 71, 3714–3723. doi: 10.1128/IAI.71.7.3714-3723.2003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JRM and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Guzmán-Herrador, Steiner, Alperi, González-Prieto, Roy and Llosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Bacterial Therapy of Cancer: Promises, Limitations, and Insights for Future Directions

M. Gabriela Kramer<sup>1,2\*</sup>, Martín Masner<sup>††</sup>, Fernando A. Ferreira<sup>1,2</sup> and Robert M. Hoffman<sup>3,4</sup>

<sup>1</sup> Department of Biotechnology, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay, <sup>2</sup> Laboratory of Carbohydrates and Glycoconjugates, Department of Organic Chemistry, Facultad de Química, Universidad de la República, Montevideo, Uruguay, <sup>3</sup> AntiCancer, Inc., San Diego, CA, United States, <sup>4</sup> Department of Surgery, University of California, San Diego, San Diego, CA, United States

## OPEN ACCESS

### Edited by:

Chew Chieng Yeo,  
Universiti Sultan Zainal Abidin,  
Malaysia

### Reviewed by:

Jung-Joon Min,  
Chonnam National University,  
South Korea  
Kaushlendra Tripathi,  
University of Alabama at Birmingham,  
United States  
Lay-Hong Chuah,  
Monash University Malaysia, Malaysia

### \*Correspondence:

M. Gabriela Kramer  
mgkramer@higiene.edu.uy

### † Present address:

Martín Masner,  
Instituto de Investigaciones en  
Ciencias de la Salud, Consejo  
Nacional de Investigaciones  
Científicas y Técnicas, Universidad  
Nacional de Córdoba, Córdoba,  
Argentina

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 31 August 2017

**Accepted:** 05 January 2018

**Published:** 23 January 2018

### Citation:

Kramer MG, Masner M, Ferreira FA  
and Hoffman RM (2018) Bacterial  
Therapy of Cancer: Promises,  
Limitations, and Insights for Future  
Directions. *Front. Microbiol.* 9:16.  
doi: 10.3389/fmicb.2018.00016

Spontaneous tumors regression has been associated with microbial infection for 100s of years and inspired the use of bacteria for anticancer therapy. Dr. William B. Coley (1862–1936), a bone- sarcoma surgeon, was a pioneer in treating his patients with both live bacterial-based and mixture of heat-killed bacteria known as “Coley’s toxins.” Unfortunately, Coley was forced to stop his work which interrupted this field for about half a century. Currently, several species of bacteria are being developed against cancer. The bacterial species, their genetic background and their infectious behavior within the tumor microenvironment are thought to be relevant factors in determining their anti-tumor effectiveness *in vivo*. In this perspective article we will update the most promising results achieved using bacterial therapy (alone or combined with other strategies) in clinically-relevant animal models of cancer and critically discuss the impact of the bacterial variants, route of administration and mechanisms of bacteria-cancer-cell interaction. We will also discuss strategies to apply this information using modern mouse models, molecular biology, genetics and imaging for future bacterial therapy of cancer patients.

**Keywords:** bacterial-based therapies, Coley’s toxins, antitumor effect, immune response, bactofection, combined therapies, *Salmonella enterica* serovar Typhimurium (S. Typhimurium), animal models of cancer

## BACK TO THE CONTROVERSIAL FUTURE

The use of microorganisms, in particular live bacteria, for prophylactic vaccination and cancer therapy have been used in humans for long periods in the past and have been a matter of controversy (Payette and Davis, 2001; Hoption Cann et al., 2003). Dr. William B. Coley in the 19th century at the New York Hospital, later to become the Memorial Sloan Kettering Cancer Center (McCarthy, 2006; Hoffman, 2016a), observed and reported spontaneous tumor regression in patients with streptococcal infections (principally erysipelas, known to be caused by *Streptococcus pyogenes*). In 1891 Dr. Coley started to treat his cancer patients with streptococcal living cultures and observed that inducing a fever was crucial for tumor regression; however such a strategy also caused some fatalities (McCarthy, 2006). Coley then generated a variety of “anti-tumor vaccines” mixing heat-killed bacteria, combining *S. pyogenes* with *Serratia marcescens*. In this way he could stimulate the symptoms of an infection (for example, inflammation, chills, fever) without the risks of a bacteremia. These vaccines became known as “Coley’s toxins” and were administered to patients with sarcomas, carcinomas, lymphomas, melanomas, and mielomas. Despite the cures and



remarkable improvements obtained in patients treated with Coley's bacterial-based therapeutics (Nauts et al., 1946; Nauts, 2004; McCarthy, 2006), his boss, the renowned pathologist James Ewing, forced Dr. Coley's to end all projects involving bacteria-based treatments alleging Coley's inconsistent data and pronouncing himself in favor of radiotherapy, which rapidly took over the market of cancer therapeutics.

Sixteen different preparations of "Coley's toxins" have been used since the method was introduced in 1892, of which three were considerably more potent than the rest (particularly the Buxton's Type VI formula). However, the only preparation available in the United States since 1921 seemed to be weaker compared to the used in the early years (Nauts et al., 1946). Coley's work gradually fell out of favor and by 1962 the Food and Drug Administration (FDA) refused to acknowledge "Coley's toxins" as an approved drug, making it illegal to prescribe them outside of clinical trials. Since then, several small clinical trials have been conducted with mixed results.

To date, *Bacillus Calmette-Guerin* (BCG), is the only bacterial agent approved by the FDA and it is employed for the treatment of superficial, non-muscle invasive bladder cancer (NMIBC) since the late 1970s (Gontero et al., 2010). BCG is an attenuated strain of *Mycobacterium bovis* obtained at the Pasteur Institute in the early 1900s. Patients typically receive repeated instillations of live bacteria into the bladder. BCG is recommended as the standard of care for high-risk NMIBC and remains the most effective intravesical treatment for this disease, although the response predictor factors of BCG are unknown (Kim and Steinberg, 2001; Zlotta et al., 2009; Gontero et al., 2010).

In the last decades a resurgence of the field has taken place and contemporary investigators demonstrated the efficacy of a number live attenuated bacteria to destroy cancer cells *in vitro*, to selectively accumulate, replicate within and destroy tumors in rodents, to induce an immune-mediated anti-tumor response and to target small metastatic nodules spread in the organism and inhibit their growth (Yu et al., 2004; Adkins et al., 2012; Hoffman, 2012b). Promising results were obtained using modern methods of bacterial genetics, cancer cell and molecular biology, and *in vivo* imaging (Min et al., 2008a,b; Uchugonova et al., 2012; Hoffman, 2015). The mechanism of action of bacterial therapy of cancer and toxicity *in vivo* is not yet clearly understood and the potential acquisition of antibiotic-resistance or mutations that would revert the bacteria attenuated phenotype could be a real risk for the patients. Therefore, the building of a broad integrated picture requires a critical scientific and medical vision, for moving forward.

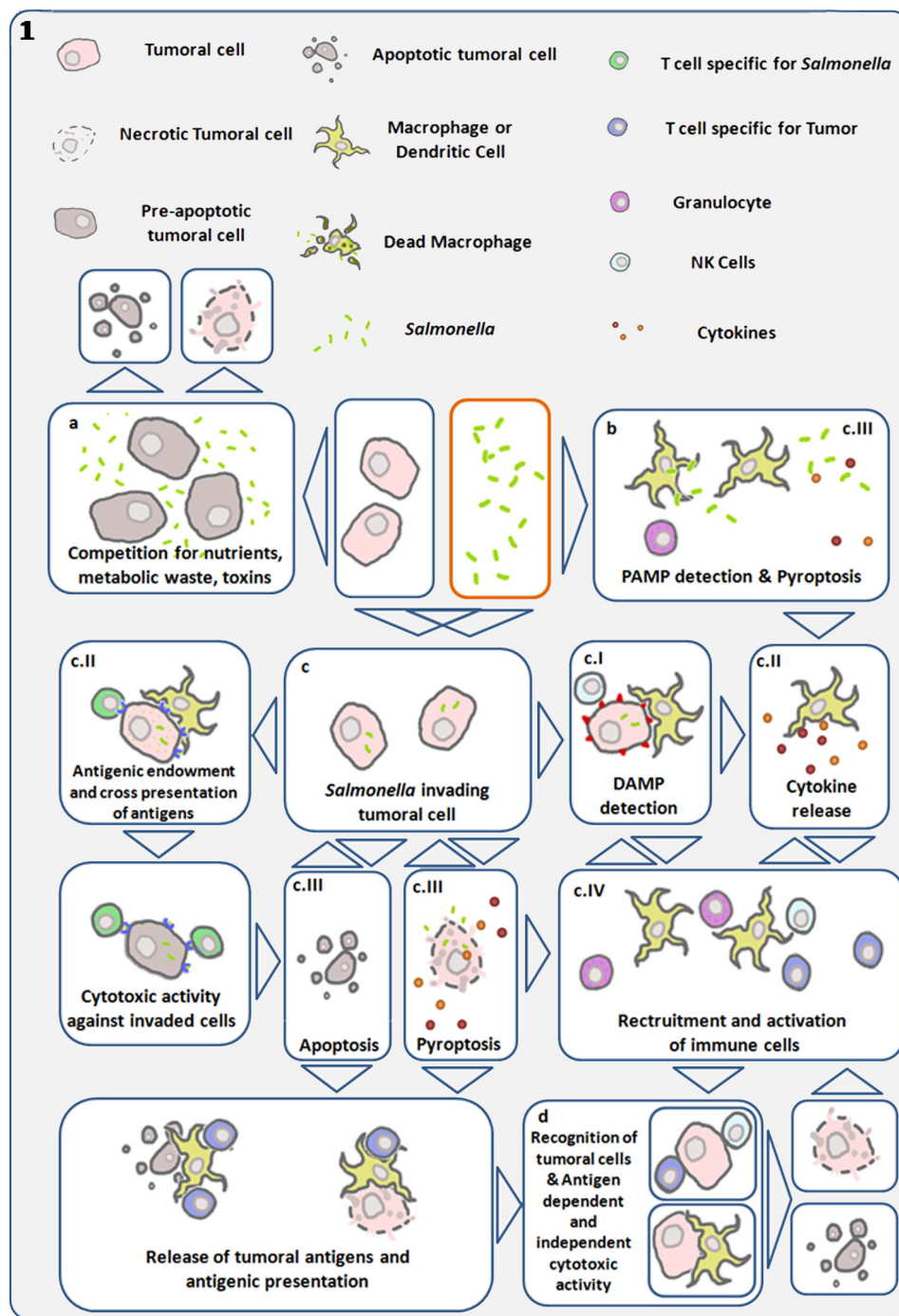
## HIGHLIGHTS BUT STILL MANY QUESTIONS

Bacteria display a number of different characteristics that could be relevant in the therapy against cancer. The direct and immune-mediated anticancer properties derive from biological interactions between the bacteria and the host tumor microenvironment. Important features of the bacteria such

as motility, tumor chemotaxis, invasive capacity, cytotoxic potential, pathogen-associated molecular patterns (PAMP) composition/abundance, among others, vary between strains and may affect how they trigger the anti-tumor response (Dang et al., 2001; Cheadle and Jackson, 2002; Hoffman, 2011; Adkins et al., 2012; Kim et al., 2015; Phan et al., 2015). Although the mechanism of bacterial tumor tropism is poorly understood there is evidence indicating that irregular organization of blood vessels within the tumor tissue that often leads to the development of hypoxic and/or necrotic regions and/or an immune-suppressive microenvironment inside the tumor mass may facilitate survival and growth of attenuated auxotrophic bacteria by providing them with nutrients and immune-protection (Forbes et al., 2003; Wouters et al., 2003; Yu et al., 2004). Moreover, niche-specific genes involved in the process of preferential tumor colonization after systemic bacteria delivery, were also identified (Silva-Valenzuela et al., 2014).

Different variants from the genera *Bifidobacterium*, *Clostridium*, *Lactococcus*, *Shigella*, *Vibrio*, *Listeria*, *Escherichia*, and *Salmonella* have been assayed in animal models of cancer (Yazawa et al., 2000; Cheadle and Jackson, 2002; Oelschlaeger, 2010; Patyar et al., 2010; Hoffman, 2012b). Obligate anaerobes such as *Bifidobacterium longum* and a *Clostridium novyi* strain devoid of its lethal toxin (*C. novyi*-NT) have shown preferential localization in low oxygenated necrotic areas of implanted tumors in mice after systemic administration, inducing tumor regression in some cases, although they were unable to grow in viable tumor tissue due to high oxygen tension, a fact that may have limited their efficacy as mono-therapy (Dang et al., 2001; Hoffman, 2012a). However, intra-tumor (i.t.) administration of *C. novyi*-NT has shown objective responses in canine tumors, which are more like those of humans because they are naturally occurring in animals with heterogeneous genetic backgrounds (Roberts et al., 2014). On the other hand, attenuated auxotrophic mutants of the facultative anaerobe *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) have been shown to invade and destroy a broad number of cancer cell types *in vitro*, as well as to replicate in hypoxic and oxidic tumor regions *in vivo*, being the most efficient anti-tumor bacteria assayed in experimental models of cancer thus far (Pawelek et al., 1997; Leschner and Weiss, 2010; Nguyen et al., 2010; Hoffman, 2011, 2016b,c). Among them, *S. Typhimurium* VNP20009, attenuated by the lipid A (*msbB*) deletion and purine (*purI*) auxotrophic mutations, has shown anti-tumor efficacy in mice and swine and was safely administrated to patients with metastatic melanoma and renal carcinoma in a Phase I clinical trial; however, efficacy was not observed, perhaps due to over-attenuation (Toso et al., 2002).

A more tumor-virulent variant and less toxic against the host is *S. Typhimurium* A1-R (Zhao et al., 2006). Unlike VNP20009, the A1-R variant was obtained by successive passages from re-infected human tumor xenografts in nude mice treated with the *S. Typhimurium* A-1 auxotrophic (Leu- Arg-dependent) parental bacteria (Zhao et al., 2005). This selection procedure may account for A1-R's particular tumor-specificity and stronger anti-tumor activity (Zhao et al., 2006). A comparative study between VNP20009 and A1-R in nude mice showed that mice



**FIGURE 1 |** Diagram showing main antitumor mechanisms induced by *S. Typhimurium* (*Salmonella*). Links are established between direct cytotoxicity induced by bacteria and indirect tumor cell death triggered by the immune system. (a) Bacterial infection within the tumor microenvironment results in inhibition of tumor growth and cell death. (b) Detection of bacterial pathogen-associated molecular patterns (PAMP) by immune cells, trigger cytokine release and recruitment of leukocytes capable of initiating anti-tumor immune responses (Patyar et al., 2010). (c) Using their Type III secretion system, *S. Typhimurium* can introduce bacterial factors in cancer cells allowing its internalization and intra-cellular replication (Avogadri et al., 2005; Knodler et al., 2010). (cl) Invasive *Salmonella* induces cell stress responses through danger-associated molecular patterns (DAMP), which are interpreted as damage signals by the immune system. (cII) Simultaneously, this same process can lead to cytokine expression and the transfer of antigens from the bacteria to the cancer cell, enabling the adaptive immune system to recognize and target the invaded cancer cell as infected and bearer of exogenous antigens (Avogadri et al., 2005). Gap junctions are concomitantly induced in the invaded cell and enable cross presentation of antigens to antigen presenting cells (Saccheri et al., 2010). Both processes can give rise to antigen-dependent elimination of infected cancer cells. (cIII) *Salmonella* can lead to the death of the infected cell, by inducing apoptosis or pyroptosis. The later is a programmed inflammatory cell death, characterized

(Continued)

**FIGURE 1 | Continued**

by activation of caspase 1, activation of the inflammasome, and IL-1B and IL-18 secretion, as well as cell rounding and detachment, cytoskeleton reorganization, nucleus deformation and rupture of the cell membrane, resulting in the release of inflammatory signals (Fink and Cookson, 2005, 2007; Knodler et al., 2010; Wang et al., 2013). This mechanism can result in cancer-cell death and immune-cell activation. Pyroptosis was first described in macrophages, which die quickly as a result of this process, and is of particular interest in cancer immunotherapy, as tumor-associated macrophages have been shown to have immune-suppressive properties. Reducing their number could be another component of the *S. Typhimurium* anti-tumor effect. Cancer cell death leads to tumor-antigen liberation, and the released bacteria can infect surrounding cancer cells. (cIV) In the process of pyroptosis, pro-inflammatory cytokines IL-1-B and IL-18 can trigger recruitment and activation of immune cells (Knodler et al., 2010; Zhao et al., 2012; Wang et al., 2013). (d) Various mechanisms enhance and converge to enable tumor-antigen recognition and activation of cytotoxic responses both in an antigen-dependent and -independent manner. *S. Typhimurium* proteins injected into the cancer cell cytosol are subject to proteasomal degradation, resulting in bacterial peptides that can be presented through MHC I to cytotoxic lymphocytes (Avogadri et al., 2005; Saccheri et al., 2010).

tolerated *S. Typhimurium* A1-R to at a least twofold higher dose than VNP20009 when the bacteria were administered intravenously (i.v.). In addition, A1-R showed higher tumor targeting and inhibited the Lewis lung carcinoma to a greater extent than VNP20009, with less body weight loss (Zhang et al., 2015). In addition, *S. Typhimurium* A1-R mono-therapy has shown to be effective against primary and metastatic human prostate, breast, and pancreatic cancer as well as osteosarcoma, fibrosarcoma, and glioma in clinically-relevant mouse models (Hoffman, 2016c and references therein). Tumors with a high degree of vascularity were more sensitive to A1-R and vascular destruction appears to play a role in A1-R anti-tumor efficacy (Liu et al., 2010). In addition, A1-R was shown to induce stem-like and non-stem cancer-cell death *in vivo*, indicating that A1-R could be used to kill chemo-resistant cancer stem-like cells (Hiroshima et al., 2013). Together these results suggest that *S. Typhimurium* A1-R may have a greater clinical potential than VNP20009 (Zhang et al., 2015) and that not only the bacterial species, but also their genetic background needs to be taken into account when searching for improvements in bacteria-based therapies.

*Salmonella Typhimurium* defective in the synthesis of ppGpp ( $\Delta$ ppGpp: depletion of *relA* and *spoT*), showed  $10^5$  to  $10^6$ -fold attenuation compared with WT strain (Na et al., 2006). This attenuated strain showed very high tumor targeting and stimulation of regional tumor immunity (Kim et al., 2015; Phan et al., 2015; Zheng et al., 2017).

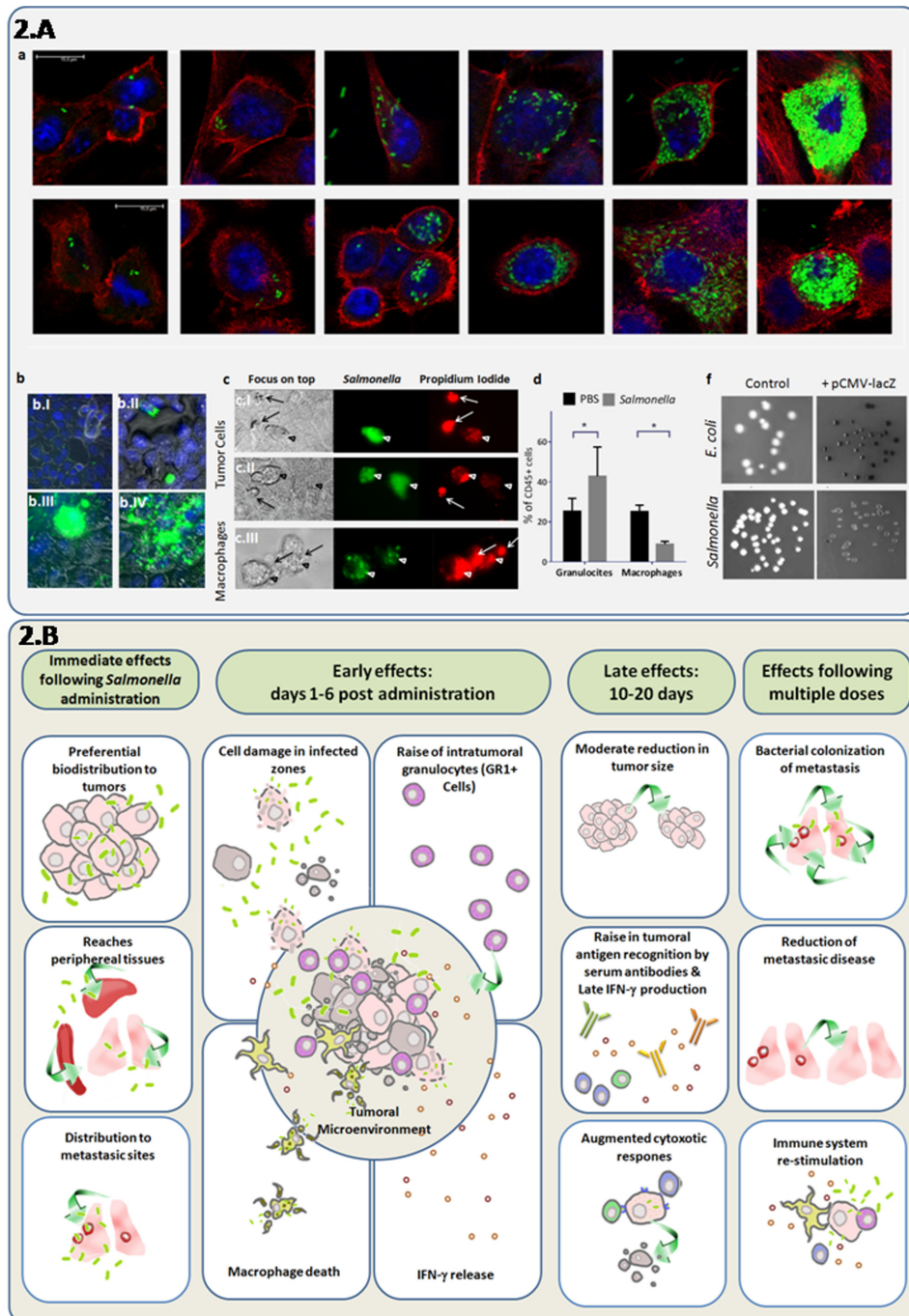
In this regard, high-throughput screenings for *Salmonella* avirulent mutants can identify variants with reduced fitness in normal tissues but unchanged fitness in tumors for potential use as cancer therapeutics (Arrach et al., 2010). As an example, a reported genetically-engineered *S. Typhimurium aroA aroD* double mutant harboring the Flt3 Ligand, used to treat melanoma in mice resulted in 50% tumor regression (Yoon et al., 2007). However, *aroA* and *aroD* were later identified by Arrach et al. (2010) as Class 2 mutants which show reduced fitness in tumors compared to Class 1 mutants, increasing the probability that a different avirulent mutant that grows better in tumors might have resulted in a more complete anti-tumor response. In a competitive fitness assay in human prostate tumors growing in mice, Class 1 mutant *STM3120* not only had a fitness advantage over Class 2 mutants, but also effectively targeted tumors after intragastric delivery, suggesting an oral route as an option for bacterial cancer therapy (Arrach et al., 2010). The ability to screen thousands of candidates and evaluate individual mutants in parallel using high-throughput sequencing offers a clear

advantage over conventional screening methods. Mutants that retain tumor-targeting while being poor colonizers of normal tissue, are desirable for cancer therapeutics.

The patient-derived xenograft (PDX) mouse models of cancer are emerging as an important component of personalized cancer therapy (Cho et al., 2016). PDX models are generated by implanting sectioned patient tumor fragments into immunodeficient mice, subcutaneously or orthotopically (into the organ or tissue of the cancer origin). Patient-derived orthotopic xenografts (PDOX) have the additional advantage that they usually metastasize as in the patient (Hiroshima et al., 2016). These models retain the histologic characteristics, heterogeneity of cancer cells and genomic signature of the patient tumor enabling the identification of effective individualized therapy (Cho et al., 2016). *S. Typhimurium* A1-R has shown to be effective against osteosarcoma in a PDX model (Murakami et al., 2017) and soft-tissue sarcoma, pancreatic cancer and melanoma in PDOX models (Hiroshima et al., 2014; Murakami et al., 2016; Yamamoto et al., 2016). Although these models need to be immunocompromised in order to allow human tumor engraftments and therefore do not allow evaluation of the immune-mediated bacterial activity, we believe that studies that employ PDOX models would allow the selection of the best-suited bacteria for individual tumors and prediction for its effectiveness in patients. "Humanized" PDOX models (Zitvogel et al., 2016) will be used to determine tumor-immunology effects of bacteria.

**Figures 1, 2** show some of the complex net of events that are involved in promoting bacterial anti-tumor efficacy. However, in most models bacteria mono-therapies are not sufficient to eliminate a primary tumor or the metastatic burden. Combined therapies including chemotherapy (Dang et al., 2001; Yamamoto et al., 2016; Yano et al., 2016), radiotherapy (Jiang et al., 2010), traditional herbal medicine (Zhang et al., 2013), anti-angiogenic and/or immunotherapy (Binder et al., 2013; Kramer et al., 2015) or the use of bacteria carrying plasmids coding for anti-tumor genes (reviewed in Moreno et al., 2010; Nguyen and Min, 2017) have shown enhanced results. Based on the use of eukaryotic gene-expression systems it has been suggested that bacteria can act as vector systems for plasmid transfer to mammalian cancer cells, a process known as "bactofection" (Weiss and Chakraborty, 2001; Baban et al., 2010; Othman et al., 2013). However, this trans-kingdom gene delivery assumption is still a matter of controversy (Gahan et al., 2009 and **Figure 2A**). Therefore, for the best performance of a bacteria + plasmid combination, the determination of





**FIGURE 2 |** Direct and synergistic anti-tumor effects of attenuated *S. Typhimurium* integrating cellular and systemic immune responses. **(A)** Induction of cell death and granulocyte recruitment associated with intracellular replication of attenuated *S. Typhimurium* LVRO1 (*Salmonella*), which previously showed a modest antitumor effect in the 4T1 metastatic breast cancer model (Kramer et al., 2015). (a) Confocal microscopy indicates bacteria invasion and replication in breast cancer cell lines: 4T1 (ATCC-CRL2539) (upper line) and NMU (ATCC-CRL1743) (lower line) in a time-course experiment. Cell cultures were grown in glass coverslips, infected with *Salmonella* expressing the GFP gene and sampled at 2, 12, 24, or 48 h to follow progression of intracellular replication. Specimens were fixed in paraformaldehyde 4%, washed in PBS and stained with DAPI and Phalloidin-Alexa555 (Invitrogen<sup>TM</sup>). After the staining, the coverslips were washed with PBS, mounted using Pro Long Gold (Invitrogen<sup>TM</sup>) and sealed with nail polish. This three color fluorescence pattern allowed the 3D analysis of the infected cultures, by simultaneously visualizing the bacteria, the nucleus and the F-actin cytoskeleton. Intracellular/extracellular determination of bacteria was possible due to the delimited borders of the actin cytoskeleton which are close to the cell membrane. Images were obtained with a LEICA<sup>®</sup>TCS SP5 II spectral confocal microscope and processed with the software Leica<sup>®</sup>LAS AF. As observed, bacterial invasion progresses, showing intracellular cytoplasmic hyperreplication over time. (b) Epifluorescence microscopy of 4T1 monolayers infected cells. Cancer cells were infected with *Salmonella*-GFP for 2 h and observed at different time points. Specimens were washed in PBS, fixed (Continued)



**FIGURE 2 | Continued**

in paraformaldehyde 4%, and stained with DAPI (Invitrogen<sup>TM</sup>). After 5 min staining, invaded cultures were washed and observed in a Nikon<sup>®</sup> Ti-S epifluorescence inverted microscope. At 2 h few peri-nuclear bacteria could be seen (b.I) At 24 h (b.II) bacteria replicated in the cytoplasm and some infected cells appear rounded and extruded. At 48 h (b.III) densely-infected cells were similar, and eventually burst and release their cellular contents (b.IV). (c) Live infected cultures were observed either intact or in the presence of propidium iodide (500 nM) to assess intracellular bacterial mobility and cell viability, respectively. Monolayers of mammary cancer cells: 4T1 (c.I) and NMU (c.II), as well as macrophage cells J774.A (c.III) were infected with *Salmonella*-GFP. At 24 h post-infection, infected cells (green) die as indicated by propidium iodide staining (red). Macrophages died at earlier time points (2–16 h). Arrows point the extruded cells. (d) Flow cytometry of intratumor immune cells at 6 days after *Salmonella* inoculation of 4T1 tumors *in vivo*. As observed, the intra-tumor granulocyte/myeloid-derived-suppressor cell (Ly6G+CD11b+) levels increase and macrophage (F4/80+CD11b+) levels decreased after bacteria administration among total leukocytes (CD45+ cells). (f) X-gal agar plates were used to seed the untransformed bacteria (control) or bacteria transformed with a plasmid containing the *lacZ* gene under the control of the eukaryotic cytomegalovirus (CMV) promoter (pCMV-lacZ). As observed, the *lacZ* gene product  $\beta$ -galactosidase was detected, indicating that the CMV promoter was active in prokaryotic cell species. **(B)** *In vivo* effects of attenuated *S. Typhimurium* (*Salmonella*) in mice bearing metastatic cancer. This integrative diagram shows the anti-tumor effects of attenuated variants of *Salmonella* evaluated as mono-therapy. The bacteria inoculation by different routes (systemic or intratumoral) results in its biodistribution to most organs, but with a marked preference for tumors, including metastatic sites (Pawelek et al., 1997; Low et al., 1999; Forbes et al., 2003; Yu et al., 2004; Hoffman, 2016a). In tumors, bacterial infection is associated with tumor-tissue architecture deterioration, a rise in granulocytic cells and INF- $\gamma$  induction and a decrease of intra-tumor macrophages (Avogadri et al., 2005; Westphal et al., 2008; Zheng et al., 2017). Late effects (10–20 days after bacteria administration) are characterized by a moderate decrease in tumor size, adaptive immune responses including INF- $\gamma$  production, antibody recognition of tumor antigens, and cytotoxic immune activities (Avogadri et al., 2005; Kramer et al., 2015; Masner et al., unpublished results). Repeated administration of attenuated bacteria could result in a better targeting of metastases (Zhao et al., 2012), while stimulating immune responses that enhance cancer-cell elimination.

the actual location of transgene expression would allow the right selection of the gene, promoter, and secretion system (if required) to achieve optimized therapy (Forbes, 2010; Zheng et al., 2017). In addition, since the bacteria usually induce death of infected cells within few hours, the rational to use bacteria as a gene delivery system (vector) to immune and/or tumor cells needs to be re-evaluated if medium- or long-term persistence of therapeutic gene expression is necessary *in vivo*.

In terms of combined therapies, a remarkable example of a neoadjuvant (pre-operative) synergistic efficacy was observed using *S. Typhimurium aro C* mutant LVR01 in combination with interleukin 12 (IL-12) expressed from the alfaviral eukaryotic gene vector SFV-IL-12 (Kramer et al., 2015). This approach was evaluated in an immunocompetent mouse model of locally-advanced breast cancer and resulted in a highly effective anti-metastatic therapy, leading to 90% disease free mice, while either mono-therapy was not effective. Moreover, the efficacy of this combined therapy depended on the order in which both agents were administered (Kramer et al., 2015). An initial anti-angiogenic effect associated with a T helper-cell-1-primed response that was timely induced seemed to account for the main global effect. However, the underlying mechanisms of this combination and timing of both factors raised various questions that remain un-answered.

Other relevant questions to be answered for bacteria-based cancer therapy optimization are related to the dose, schedule, and route of administration. A dose-dependent effect of attenuated *S. Typhimurium* was observed, as well as multiple dosing are more efficient than mono-doses (Hayashi et al., 2009; Nagakura et al., 2009; Grille et al., 2014), although the range needs to be determined to avoid toxicity (Zhao et al., 2012). The efficacy and safety of three different routes of *S. Typhimurium* A1-R administration: oral (p.o.), i.v. and i.t. was compared in nude mice with orthotopic human breast cancer indicated that the p.o. route was safer, and the i.v. route was more effective (Zhang et al., 2012). However, such experiments may need to be performed for each type of tumor, since it was also shown in a model of disseminated human ovarian cancer treated with i.v. and

intraperitoneal (i.p.) *S. Typhimurium* A1-R, that i.p. treatment was less toxic than i.v. administration (Matsumoto et al., 2015).

Although useful in many approaches, human xenografted tumors into immunodeficient mice limit our knowledge about the range of effects that certain bacterial strains can exert. In this regard, studies in immunocompetent animals are more representative of the complex spectrum of interactions between the bacteria and the tumor microenvironment, thereby enabling immune effects that are otherwise absent in immunocompromised mice. This could be crucial for “tuning” the bacteria to the right degree of immunogenicity/attenuation, avoiding shock while promoting adjuvant effects (Yu et al., 2004). Moreover, toxicity issues regarding immunotherapies are a main concern today. From acute shock to autoimmune diseases, we could gain a better understanding of the risk of side effects of bacteria therapy of cancer from preclinical models that include all the functional branches of the immune system. Undesirable attenuated bacterial infection can be in theory treated with antibiotics; however, long-term clinical trials in humans are required to evaluate toxicity in detail, since the chance of septic shock and/or tumor lysis syndrome could be a fact. In addition, we believe that, there is still a considerable need of work to evaluate bacteria for natural acquisition of antibiotic-resistant genes and/or reversion of attenuation mutations, as well as comparing the anti-tumor efficacy and secondary effects of bacteria or bacterial products *versus* conventional therapies. Moreover, we cannot rule out the possible clearance of bacteria by the immune system before reaching the tumor site in a patient-dependant manner, resulting in treatment failure.

## THE FUTURE OF WHAT Dr. Coley BEGAN

Each of the 16 “Coley’s toxins” that have been used might have a complex and variable composition, including components of the culture media, products released by the bacteria in the medium, components relevant by bacteria lysis (and autolysis). The inactivation method used to prepare the vaccine and the inclusion, or not, of a filtration step in the preparation of the

toxins will affect the final products. The i.v. administration of a suspension of inactivated bacteria cells may mimic a nanodrug, and the number of particles, their size, shape, charge, and surface molecules may affect the immune system response (van Riet et al., 2014).

Both *Streptococcus pyogenes* and *Serratia marcescens* produce exotoxins. *S. pyogenes* produces the pyrogenic exotoxins SpeA, SpeB, and SpeC which have the capacity to unspecifically stimulate CD4+ lymphocytes, leading to a strong secretion of different cytokines (Babbar, 2015). *S. marcescens*, produces prodigiosin, a low-molecular weight red pigmented heterocyclic tripyrrolic toxin with anti-tumor activity (Elahian et al., 2013). The toxins, together with other components of the formulation, result in generation of fever and potential anti-tumor response. The administration route may also influence the efficacy of Coley's toxins including i.v., i.p., direct injection in the tumor, or subcutaneous or intramuscular administration (Nauts et al., 1946).

A chemical description of "Coley's toxins" can be assessed using the analytical tools currently used for proteomic and metabolomic studies (Wishar, 2016). Nuclear magnetic resonance (NMR) and mass spectrometry (MS) methods for the analysis of high- and low-molecular weight components of complex mixtures or their derivatives (Alonso et al., 2015) could be used in combination with multivariate analysis to identify the components responsible for anti-tumor activity. The identification of the active components and their mode of action, would allow the selection of more active and better-defined vaccines, as well as the design of tailored formulations capable of producing the right amount of systemic or tumor-localized fever (Noe, 2016) for optimal stimulation of the host immune system and cytokine secretion to achieve best anti-tumor efficacy.

## REFERENCES

- Adkins, I., Holubova, J., Kosova, M., and Sadilkova, L. (2012). Bacteria and their toxins tamed for immunotherapy. *Curr. Pharm. Biotechnol.* 13, 1446–1473. doi: 10.2174/138920112800784835
- Alonso, A., Marsal, S., and Julià, A. (2015). Analytical methods in untargeted metabolomics: state of the art in 2015. *Front. Bioeng. Biotechnol.* 3:23. doi: 10.3389/fbioe.2015.00023
- Arrach, N., Cheng, P., Zhao, M., Santiviago, C. A., Hoffman, R. M., and McClelland, M. (2010). High-throughput screening for *Salmonella* avirulent mutants that retain targeting of solid tumors. *Cancer Res.* 70, 2165–2170. doi: 10.1158/0008-5472.CAN-09-4005
- Avogadri, F., Martinoli, C., Petrovska, L., Chiodoni, C., Transidico, P., Bronte, V., et al. (2005). Cancer immunotherapy based on killing of *Salmonella*-infected tumor cells. *Cancer Res.* 65, 3920–3927. doi: 10.1158/0008-5472.CAN-04-3002
- Babbar, C. K., Cronin, M., O'Hanlon, D., O'Sullivan, G. C., and Tangney, M. (2010). Bacteria as vectors for gene therapy of cancer. *Bioeng. Bugs* 1, 385–394. doi: 10.4161/bbug.1.6.13146
- Babbar, A. (2015). *Streptococcal Superantigens*, SpringerBriefs in Microbiology. Cham: Springer, 1–41. doi: 10.1007/978-3-319-22455-8\_1
- Binder, D. C., Engels, B., Arina, A., Yu, P., Schlauch, J. M., Fu, Y.-X., et al. (2013). Antigen-specific bacterial vaccine combined with anti-PD-L1 rescues dysfunctional endogenous T cells to reject long-established cancer. *Cancer Immunol. Res.* 1, 123–133. doi: 10.1158/2326-6066.CIR-13-0058
- Cheadle, E. J., and Jackson, A. M. (2002). Bugs as drugs for cancer. *Immunology* 107, 10–19. doi: 10.1046/j.1365-2567.2002.01498.x

## AUTHOR CONTRIBUTIONS

MGK received the invitation to contribute in this special issue, designed and wrote most of the sections and supervised the artwork. MM made all the drawings, acquired and analyzed the data and participated in the text writing. FAF wrote the final section and was essential in motivating the team work. RMH inspired the main ideas of this perspective article, supplied relevant literature and critically revised the manuscript.

## FUNDING

Grant support was from Agencia Nacional de Investigación e Innovación (ANII) and Comisión Honoraria de Lucha contra el Cáncer (CHLCC) to MGK. MM was recipient of an ANII postgraduate studentship.

## ACKNOWLEDGMENTS

The authors thank Dr. Fernando Gonzalez (Department of Biophysics, Udelar) for assistance with confocal microscopy and image processing; Dr. Patricia Berasain (Parasitology Unit, Udelar) for help with informatics; Dr. Teresa Freire (Department of Immunobiology, Udelar) for NMU cells; Dr. Lucia Veiga and Dr. José A. Chabalgoity (Department of Biotechnology, Udelar) for LVR01-GFP; Dr. Helen Bauer (German Research Centre for Biotechnology, Germany) for pCMV-lacZ and Rodrigo Gonzalez (undergraduate student supervised by MGK) for X-gal staining of transformed bacteria.

- Cho, S.-Y., Kang, W., Han, J. Y., Min, S., Kang, J., Lee, A., et al. (2016). An integrative approach to precision cancer medicine using patient-derived xenografts. *Mol. Cells* 39, 77–86. doi: 10.14348/molcells.2016.2350
- Dang, L. H., Bettgowda, C., Huso, D. L., Kinzler, K. W., and Vogelstein, B. (2001). Combination bacteriolytic therapy for the treatment of experimental tumors. *Proc. Natl. Acad. Sci. U.S.A.* 98, 15155–15160. doi: 10.1073/pnas.251543698
- Elahian, F., Moghimi, B., Dinmohammadi, F., Ghamghami, M., Hamidi, M., and Mirzaei, S. A. (2013). The anticancer agent prodigiosin is not a multidrug resistance protein substrate. *DNA Cell Biol.* 32, 90–97. doi: 10.1089/dna.2012.1902
- Fink, S. L., and Cookson, B. T. (2005). Apoptosis, pyroptosis, and necrosis: mechanistic description of dead and dying eukaryotic cells. *Infect. Immun.* 73, 1907–1916. doi: 10.1128/IAI.73.4.1907
- Fink, S. L., and Cookson, B. T. (2007). Pyroptosis and host cell death responses during *Salmonella* infection. *Cell. Microbiol.* 9, 2562–2570. doi: 10.1111/j.1462-5822.2007.01036.x
- Forbes, N. S. (2010). Engineering the perfect (bacterial) cancer therapy. *Nat. Rev. Cancer* 10, 785–794. doi: 10.1038/nrc2934
- Forbes, N. S., Munn, L. L., Fukumura, D., and Jain, R. K. (2003). Sparse initial entrapment of systemically injected *Salmonella typhimurium* leads to heterogeneous accumulation within tumors. *Cancer Res.* 63, 5188–5193.
- Gahan, M. E., Webster, D. E., Wesselingh, S. L., Strugnell, R. A., and Yang, J. (2009). Bacterial antigen expression is an important component in inducing an immune response to orally administered *Salmonella*-delivered DNA vaccines. *PLOS ONE* 4:e6062. doi: 10.1371/journal.pone.0006062
- Gontero, P., Bohle, A., Malmstrom, P.-U., O'Donnell, M. A., Oderda, M., Sylvester, R., et al. (2010). The role of bacillus Calmette-Guérin in the treatment

- of non-muscle-invasive bladder cancer. *Eur. Urol.* 57, 410–429. doi: 10.1016/j.eururo.2009.11.023
- Grille, S., Moreno, M., Bascuas, T., Marqués, J. M., Muñoz, N., Lens, D., et al. (2014). *Salmonella enterica* serovar Typhimurium immunotherapy for B-cell lymphoma induces broad anti-tumour immunity with therapeutic effect. *Immunology* 143, 428–437. doi: 10.1111/imm.12320
- Hayashi, K., Zhao, M., Yamauchi, K., Yamamoto, N., Tsuchiya, H., Tomita, K., et al. (2009). Systemic targeting of primary bone tumor and lung metastasis of high-grade osteosarcoma in nude mice with a tumor-selective strain of *Salmonella typhimurium*. *Cell Cycle* 8, 870–875. doi: 10.4161/cc.8.6.7891
- Hiroshima, Y., Maawy, A., Zhang, Y., Zhang, N., Murakami, T., Chishima, T., et al. (2016). Patient-derived mouse models of cancer need to be orthotopic in order to evaluate targeted anti-metastatic therapy. *Oncotarget* 7, 71696–71702. doi: 10.18632/oncotarget.12322
- Hiroshima, Y., Zhao, M., Maawy, A., Zhang, Y., Katz, M. H. G., Fleming, J. B., et al. (2014). Efficacy of *Salmonella typhimurium* A1-R versus chemotherapy on a pancreatic cancer patient-derived orthotopic xenograft (PDOX). *J. Cell. Biochem.* 115, 1254–1261. doi: 10.1002/jcb.24769
- Hiroshima, Y., Zhao, M., Zhang, Y., Maawy, A., Hassanein, M. K., Uehara, F., et al. (2013). Comparison of efficacy of *Salmonella typhimurium* A1-R and chemotherapy on stem-like and non-stem human pancreatic cancer cells. *Cell Cycle* 12, 2774–2780. doi: 10.4161/cc.25872
- Hoffman, R. M. (2011). Tumor-seeking *Salmonella* amino acid auxotrophs. *Curr. Opin. Biotechnol.* 22, 917–923. doi: 10.1016/j.copbio.2011.03.009
- Hoffman, R. M. (2012a). Bugging tumors. *Cancer Discov.* 2, 588–590. doi: 10.1158/2159-8290.CD-12-0227
- Hoffman, R. M. (2012b). The preclinical discovery of bacterial therapy for the treatment of metastatic cancer with unique advantages. *Expert Opin. Drug Discov.* 7, 73–83. doi: 10.1517/17460441.2012.644534
- Hoffman, R. M. (2015). “Back to the future: Are tumor-targeting bacteria the next-generation Cancer therapy?,” in *Gene Therapy of Solid Cancers. Methods in Molecular Biology*, Vol. 1317, eds W. Walther and U. Stein (New York, NY: Humana Press), 239–260. doi: 10.1007/978-1-4939-2727-2\_14
- Hoffman, R. M. (ed.) (2016a). “Preface: a brief history of bacterial therapy of cancer,” in *Methods in Molecular Biology*, (New York, NY: Humana Press), doi: 10.1007/978-1-4939-3515-4
- Hoffman, R. M. (ed.) (2016b). “Tumor-Targeting *Salmonella typhimurium* A1-R: an Overview,” in *Methods in Molecular Biology*, (New York, NY: Humana Press), 1–8. doi: 10.1007/978-1-4939-3515-4\_1
- Hoffman, R. M. (ed.) (2016c). “Future of bacterial therapy of cancer,” in *Methods in Molecular Biology*, (New York, NY: Humana Press), 177–184. doi: 10.1007/978-1-4939-3515-4\_15
- Hoption Cann, S. A., van Netten, J. P., and van Netten, C. (2003). Dr. William Coley and tumour regression: a place in history or in the future. *Postgrad. Med. J.* 79, 672–680.
- Jiang, S. N., Phan, T. X., Nam, T. K., Nguyen, V. H., Kim, H. S., Bom, H. S., et al. (2010). Inhibition of tumor growth and metastasis by a combination of *Escherichia coli*-mediated cytolytic therapy and radiotherapy. *Mol. Ther.* 18, 635–642. doi: 10.1038/mt.2009.295
- Kim, J. C., and Steinberg, G. (2001). The limits of bacillus Calmette-Guerin for carcinoma *in situ* of the bladder. *J. Urol.* 165, 745–756. doi: 10.1016/S0022-5347(05)66518-4
- Kim, J. E., Phan, T. X., Nguyen, V. H., Dinh-Vu, H. V., Zheng, J. H., Yun, M., et al. (2015). *Salmonella typhimurium* suppresses tumor growth via the pro-inflammatory cytokine interleukin-1 $\beta$ . *Theranostics* 5, 1328–1342. doi: 10.7150/thno.11432
- Knodler, L. A., Vallance, B. A., Celli, J., Winfree, S., Hansen, B., Montero, M., et al. (2010). Dissemination of invasive *Salmonella* via bacterial-induced extrusion of mucosal epithelia. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17733–17738. doi: 10.1073/pnas.1006098107
- Kramer, M. G., Masner, M., Casales, E., Moreno, M., Smerdou, C., and Chabalgoity, J. A. (2015). Neoadjuvant administration of Semliki Forest virus expressing interleukin-12 combined with attenuated *Salmonella* eradicates breast cancer metastasis and achieves long-term survival in immunocompetent mice. *BMC Cancer* 15:620. doi: 10.1186/s12885-015-1618-x
- Leschner, S., and Weiss, S. (2010). *Salmonella*-allies in the fight against cancer. *J. Mol. Med.* 88, 763–773. doi: 10.1007/s00109-010-0636-z
- Liu, F., Zhang, L., Hoffman, R. M., and Zhao, M. (2010). Vessel destruction by tumor-targeting *Salmonella typhimurium* A1-R is enhanced by high tumor vascularity. *Cell Cycle* 9, 4518–4524. doi: 10.4161/cc.9.22.13744
- Low, K. B., Ittensohn, M., Le, T., Platt, J., Sodi, S., Amoss, M., et al. (1999). Lipid A mutant *Salmonella* with suppressed virulence and TNF $\alpha$  induction retain tumor-targeting in vivo. *Nat. Biotechnol.* 17, 37–41. doi: 10.1038/5205
- Matsumoto, Y., Miwa, S., Zhang, Y., Zhao, M., Yano, S., Uehara, F., et al. (2015). Intraperitoneal administration of tumor-targeting *Salmonella typhimurium* A1-R inhibits disseminated human ovarian cancer and extends survival in nude mice. *Oncotarget* 6, 11369–11377. doi: 10.18632/oncotarget.3607
- McCarthy, E. F. (2006). The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. *Iowa Orthop. J.* 26, 154–158.
- Min, J. J., Kim, H. J., Park, J. H., Moon, S. M., Jeong, J. H., Hong, Y. J., et al. (2008a). Noninvasive real-time imaging of tumors and metastases using tumor-targeting light-emitting *Escherichia coli*. *Mol. Imaging Biol.* 10, 54–61.
- Min, J. J., Nguyen, V. H., Kim, H. J., Hong, Y., and Choy, H. E. (2008b). Quantitative bioluminescence imaging of tumor-targeting bacteria in living animals. *Nat. Protoc.* 3, 629–636. doi: 10.1038/nprot.2008.32
- Moreno, M., Kramer, M. G., Yim, L., and Chabalgoity, J. A. (2010). *Salmonella* as live trojan horse for vaccine development and cancer gene therapy. *Curr. Gene Ther.* 10, 56–76. doi: 10.2174/156652310790945566
- Murakami, T., DeLong, J., Eilber, F. C., Zhao, M., Zhang, Y., Zhang, N., et al. (2016). Tumor-targeting *Salmonella typhimurium* A1-R in combination with doxorubicin eradicate soft tissue sarcoma in a patient-derived orthotopic xenograft (PDOX) model. *Oncotarget* 7, 12783–12790. doi: 10.18632/oncotarget.7226
- Murakami, T., Igarashi, K., Kawaguchi, K., Kiyuna, T., Zhang, Y., Zhao, M., et al. (2017). Tumor-targeting *Salmonella typhimurium* A1-R regresses an osteosarcoma in a patient-derived xenograft model resistant to a molecular-targeting drug. *Oncotarget* 8, 8035–8042. doi: 10.18632/oncotarget.14040
- Na, H. S., Kim, H. J., Lee, H. C., Hong, Y., Rhee, J. H., and Choy, H. E. (2006). Immune response induced by *Salmonella typhimurium* defective in ppGpp synthesis. *Vaccine* 24, 2027–2034. doi: 10.1016/j.vaccine.2005.11.031
- Nagakura, C., Hayashi, K., Zhao, M., Yamauchi, K., Yamamoto, N., Tsuchiya, H., et al. (2009). Efficacy of a genetically-modified *Salmonella typhimurium* in an orthotopic human pancreatic cancer in nude mice. *Anticancer. Res.* 29, 1873–1878.
- Nauts, H. C. (2004). *Coley's Toxins- the First Century*. Available at: <https://www.highbeam.com/doc/1G1-117854757.html>
- Nauts, H. C., Swift, W. E., and Coley, B. L. (1946). The treatment of malignant tumors by bacterial toxins as developed by the late William B. Coley, M.D., reviewed in the light of modern research. *Cancer Res.* 6, 205–216.
- Nguyen, V. H., Kim, H. S., Ha, J. M., Hong, Y., Choy, H. E., and Min, J. J. (2010). Genetically engineered *Salmonella typhimurium* as an imageable therapeutic probe for cancer. *Cancer Res.* 70, 18–23. doi: 10.1158/0008-5472.CAN-09-3453
- Nguyen, V. H., and Min, J. J. (2017). *Salmonella*-mediated cancer therapy: roles and potential. *Nucl. Med. Mol. Imaging* 51, 118–126. doi: 10.1007/s13139-016-0415-z
- Noe, G. (2016). A hybrid fever therapy for increased tumor selectivity. *Med. Hypotheses* 89, 63–64. doi: 10.1016/j.mehy.2016.02.005
- Oelschlaeger, T. A. (2010). Bacteria as tumor therapeutics? *Bioeng. Bugs* 1, 146–147. doi: 10.4161/bbug.1.2.11248
- Othman, S., Roe, A. J., Parton, R., and Coote, J. G. (2013). Use of a dual reporter plasmid to demonstrate Bactofection with an attenuated AroA<sup>-</sup> derivative of *Pasteurella multocida* B:2. *PLOS ONE* 8:e71524. doi: 10.1371/journal.pone.0071524
- Patyar, S., Joshi, R., Byrav, D. S. P., Prakash, A., Medhi, B., and Das, B. K. (2010). Bacteria in cancer therapy: a novel experimental strategy. *J. Biomed. Sci.* 17, 21. doi: 10.1186/1423-0127-17-21
- Pawelek, J. M., Low, K. B., and Bermudes, D. (1997). Tumor-targeted *Salmonella* as a novel anticancer vector. *Cancer Res.* 57, 4537–4544.
- Payette, P. J., and Davis, H. L. (2001). History of vaccines and positioning of current trends. *Curr. Drug Targets Infect. Disord.* 1, 241–247. doi: 10.2174/1568005014606017
- Phan, T. X., Nguyen, V. H., Duong, M. T., Hong, Y., Choy, H. E., and Min, J. J. (2015). Activation of inflammasome by attenuated *Salmonella typhimurium* in bacteria-mediated cancer therapy. *Microbiol. Immunol.* 59, 664–675. doi: 10.1111/1348-0421.12333

- Roberts, N. J., Zhang, L., Janku, F., Collins, A., Bai, R.-Y., Staedtke, V., et al. (2014). Intratumoral injection of *Clostridium novyi*-NT spores induces anti-tumor responses. *Sci. Transl. Med.* 6, 249ra111. doi: 10.1126/scitranslmed.3008982
- Saccheri, F., Pozzi, C., Avogadri, F., Barozzi, S., Faretta, M., Fusi, P., et al. (2010). Bacteria-induced gap junctions in tumors favor antigen cross-presentation and antitumor immunity. *Sci. Transl. Med.* 2:44ra57. doi: 10.1126/scitranslmed.3000739
- Silva-Valenzuela, C. A., Desai, P. T., Molina-Quiroz, R. C., Pezoa, D., Zhang, Y., Porwollik, S., et al. (2014). Solid tumors provide niche-specific conditions that lead to preferential growth of *Salmonella*. *Oncotarget* 7, 35169–35180. doi: 10.18632/oncotarget.9071
- Toso, J. F., Gill, V. J., Hwu, P., Marincola, F. M., Restifo, N. P., Schwartzentruber, D. J., et al. (2002). Phase I study of the intravenous administration of attenuated *Salmonella typhimurium* to patients with metastatic melanoma. *J. Clin. Oncol.* 20, 142–152. doi: 10.1200/JCO.2002.20.1.142
- Uchugonova, A., Zhao, M., Zhang, Y., Weinigel, M., König, K., and Hoffman, R. M. (2012). Cancer-cell killing by engineered *Salmonella* imaged by multiphoton tomography in live mice. *Anticancer. Res.* 32, 4331–4337.
- van Riet, E., Ainai, A., Suzuki, T., Kersten, G., and Hasegawa, H. (2014). Combatting infectious diseases; nanotechnology as a platform for rational vaccine design. *Adv. Drug Deliv. Rev.* 74, 28–34. doi: 10.1016/j.addr.2014.05.011
- Wang, Q., Imamura, R., Motani, K., Kushiya, H., Nagata, S., and Suda, T. (2013). Pyroptotic cells externalize eat-me and release find-me signals and are efficiently engulfed by macrophages. *Int. Immunol.* 25, 363–372. doi: 10.1093/intimm/dxs161
- Weiss, S., and Chakraborty, T. (2001). Transfer of eukaryotic expression plasmids to mammalian host cells by bacterial carriers. *Curr. Opin. Biotechnol.* 12, 467–472. doi: 10.1016/S0958-1669(00)00247-0
- Westphal, K., Leshner, S., Jablonska, J., Loessner, H., and Weiss, S. (2008). Containment of tumor-colonizing bacteria by host neutrophils. *Cancer Res.* 68, 2952–2960. doi: 10.1158/0008-5472.CAN-07-2984
- Wishar, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* 15, 473–484. doi: 10.1038/nrd.2016.32
- Wouters, B. G., Koritzinsky, M., Chiu, R. K., Theys, J., Buijsen, J., and Lambin, P. (2003). Modulation of cell death in the tumor microenvironment. *Semin. Radiat. Oncol.* 13, 31–41. doi: 10.1053/srao.2003.50004
- Yamamoto, M., Zhao, M., Hiroshima, Y., Zhang, Y., Shurell, E., Eilber, F. C., et al. (2016). Efficacy of tumor-targeting *Salmonella* A1-R on a melanoma patient-derived orthotopic xenograft (PDOX) nude-mouse model. *PLOS ONE* 11:e0160882. doi: 10.1371/journal.pone.0160882
- Yano, S., Takehara, K., Zhao, M., Tan, Y., Han, Q., Li, S., et al. (2016). Tumor-specific cell-cycle decoy by *Salmonella typhimurium* A1-R combined with tumor-selective cell-cycle trap by methioninase overcome tumor intrinsic chemoresistance as visualized by FUCCI imaging. *Cell Cycle* 15, 1715–1723. doi: 10.1080/15384101.2016.1181240
- Yazawa, K., Fujimori, M., Amano, J., Kano, Y., and Taniguchi, S. (2000). Bifidobacterium longum as a delivery system for cancer gene therapy: selective localization and growth in hypoxic tumors. *Cancer Gene Ther.* 7, 269–274. doi: 10.1038/sj.cgt.7700122
- Yoon, W. S., Choi, W. C., Sin, J.-I., and Park, Y. K. (2007). Anti-tumor therapeutic effects of *Salmonella typhimurium* containing Flt<sub>3</sub> Ligand expression plasmids in melanoma-bearing mouse. *Biotechnol. Lett.* 29, 511–516. doi: 10.1007/s10529-006-9270-9
- Yu, Y. A., Shabahang, S., Timiryasova, T. M., Zhang, Q., Beltz, R., Gentschev, I., et al. (2004). Visualization of tumors and metastases in live animals with bacteria and vaccinia virus encoding light-emitting proteins. *Nat. Biotechnol.* 22, 313–320. doi: 10.1038/nbt937
- Zhang, Y., Tome, Y., Suetsugu, A., Zhang, L., Zhang, N., Hoffman, R. M., et al. (2012). Determination of the optimal route of administration of *Salmonella typhimurium* A1-R to target breast cancer in nude mice. *Anticancer. Res.* 32, 2501–2508.
- Zhang, Y., Zhang, N., Su, S., Hoffman, R. M., and Zhao, M. (2013). *Salmonella typhimurium* A1-R tumor targeting in immunocompetent mice is enhanced by a traditional Chinese medicine herbal mixture. *Anticancer. Res.* 33, 1837–1843.
- Zhang, Y., Zhang, N., Zhao, M., and Hoffman, R. M. (2015). Comparison of the selective targeting efficacy of *Salmonella typhimurium* A1-R and VNP20009 on the Lewis lung carcinoma in nude mice. *Oncotarget* 6, 14625–14631. doi: 10.18632/oncotarget.3342
- Zhao, M., Suetsugu, A., Ma, H., Zhang, L., Liu, F., Zhang, Y., et al. (2012). Efficacy against lung metastasis with a tumor-targeting mutant of *Salmonella typhimurium* in immunocompetent mice. *Cell Cycle* 11, 187–193. doi: 10.4161/cc.11.1.18667
- Zhao, M., Yang, M., Li, X.-M., Jiang, P., Baranov, E., Li, S., et al. (2005). Tumor-targeting bacterial therapy with amino acid auxotrophs of GFP-expressing *Salmonella typhimurium*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 755–760. doi: 10.1073/pnas.0408422102
- Zhao, M., Yang, M., Ma, H., Li, X., Tan, X., Li, S., et al. (2006). Targeted therapy with a *Salmonella typhimurium* leucine-arginine auxotroph cures orthotopic human breast tumors in nude mice. *Cancer Res.* 66, 7647–7652. doi: 10.1158/0008-5472.CAN-06-0716
- Zheng, J. H., Nguyen, V. H., Jiang, S.-N., Park, S.-H., Tan, W., Hong, S. H., et al. (2017). Two-step enhanced cancer immunotherapy with engineered *Salmonella typhimurium* secreting heterologous flagellin. *Sci. Transl. Med.* 9:eak9537. doi: 10.1126/scitranslmed.aak9537
- Zitvogel, L., Pitt, J. M., Daillère, R., Smyth, M. J., and Kroemer, G. (2016). Mouse models in oncoimmunology. *Nat. Rev. Cancer* 16, 759–773. doi: 10.1038/nrc.2016.91
- Zlotta, A. R., Fleshner, N. E., and Jewett, M. A. (2009). The management of BCG failure in non-muscle-invasive bladder cancer: an update. *Can. Urol. Assoc. J.* 3, S199–S205.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kramer, Masner, Ferreira and Hoffman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





## OPEN ACCESS

## Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

## Reviewed by:

Gemma Reguera,  
Michigan State University, USA  
Grzegorz Węgrzyn,  
University of Gdańsk, Poland

## \*Correspondence:

Rafael Giraldo  
rgiraldo@cib.csic.es

## †Present address:

Laura Molina-García,  
Department of Cell  
and Developmental Biology, University  
College London, UK;  
María Moreno-del Álamo,  
Department of Microbial  
Biotechnology, National Centre  
for Biotechnology, Consejo Superior  
de Investigaciones Científicas,  
Madrid, Spain;  
Zaira Martín-Moldes,  
Department of Biomedical  
Engineering, Tufts University,  
Medford, MA, USA;  
Alicia Sánchez-Gorostiaga,  
Department of Ecology  
and Evolutionary Biology, Microbial  
Sciences Institute, Yale University,  
West Haven, CT, USA;  
Aida Alonso-del Valle,  
Department of Virology  
and Microbiology, Centre  
for Molecular Biology "Severo  
Ochoa", Consejo Superior  
de Investigaciones Científicas –  
Universidad Autónoma de Madrid,  
Madrid, Spain

‡These authors have contributed  
equally to this work.

## Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 20 January 2017

Accepted: 14 March 2017

Published: 04 April 2017

# Outlining Core Pathways of Amyloid Toxicity in Bacteria with the RepA-WH1 Prionoid

Laura Molina-García<sup>1†‡</sup>, María Moreno-del Álamo<sup>1†‡</sup>, Pedro Botías<sup>2</sup>,  
Zaira Martín-Moldes<sup>3†</sup>, María Fernández<sup>4</sup>, Alicia Sánchez-Gorostiaga<sup>5†</sup>,  
Aida Alonso-del Valle<sup>1†</sup>, Juan Nogales<sup>3</sup>, Jesús García-Cantalejo<sup>2</sup> and Rafael Giraldo<sup>1\*</sup>

<sup>1</sup> Department of Cellular and Molecular Biology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Genomics Unit, Complutense University, Madrid, Spain, <sup>3</sup> Department of Environmental Biology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>4</sup> Proteomics Facility, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>5</sup> Department of Microbial Biotechnology, National Centre for Biotechnology, Consejo Superior de Investigaciones Científicas, Madrid, Spain

The synthetic bacterial prionoid RepA-WH1 causes a vertically transmissible amyloid proteinopathy in *Escherichia coli* that inhibits growth and eventually kills the cells. Recent *in vitro* studies show that RepA-WH1 builds pores through model lipid membranes, suggesting a possible mechanism for bacterial cell death. By comparing acutely (A31V) and mildly ( $\Delta$ N37) cytotoxic mutant variants of the protein, we report here that RepA-WH1(A31V) expression decreases the intracellular osmotic pressure and compromise bacterial viability under either aerobic or anaerobic conditions. Both are effects expected from threatening membrane integrity and are in agreement with findings on the impairment by RepA-WH1(A31V) of the proton motive force (PMF)-dependent transport of ions ( $\text{Fe}^{3+}$ ) and ATP synthesis. Systems approaches reveal that, in aerobiosis, the PMF-independent respiratory dehydrogenase NdhII is induced in response to the reduction in intracellular levels of iron. While NdhII is known to generate  $\text{H}_2\text{O}_2$  as a by-product of the autoxidation of its FAD cofactor, key proteins in the defense against oxidative stress (OxyR, KatE), together with other stress-resistance factors, are sequestered by co-aggregation with the RepA-WH1(A31V) amyloid. Our findings suggest a route for RepA-WH1 toxicity in bacteria: a primary hit of damage to the membrane, compromising bionergetics, triggers a stroke of oxidative stress, which is exacerbated due to the aggregation-dependent inactivation of enzymes and transcription factors that enable the cellular response to such injury. The proteinopathy caused by the prion-like protein RepA-WH1 in bacteria recapitulates some of the core hallmarks of human amyloid diseases.

**Keywords:** amyloid proteinopathy, model amyloid disease, prionoid, systems analysis, *Escherichia coli*, membrane targeting, ROS toxicity

## INTRODUCTION

Amyloids are stable and relatively simple, albeit polymorphic, structures in which peptide stretches from a given protein assemble as fibrillar  $\beta$ -sheet polymers of indefinite length (Riek and Eisenberg, 2016). The aggregation of proteins as amyloids is at the basis of many neurodegenerative and systemic human diseases (Eisenberg and Jucker, 2012). There are many

proposed routes for amyloid cytotoxicity, including the targeting of cell membranes (Butterfield and Lashuel, 2010), co-aggregation of essential cell factors (Olzscha et al., 2011; Hosp et al., 2015), interference with intracellular traffic (Woerner et al., 2016) or overloading the protein quality triage machinery, including chaperones, the proteasome and autophagy (Hipp et al., 2014). Interestingly, mitochondria, the power engines of eukaryotic cells, have recently attracted much attention due to their involvement in several amyloid proteinopathies (Lin and Beal, 2006; Liu et al., 2015). A pioneering systems biology work reported that the disease caused in mice by distinct strains of the prion protein PrP was affecting, besides other neural and glial processes, the energetic metabolism at mitochondria (Hwang et al., 2009). Later proteomic studies revealed a major presence of mitochondrial factors co-aggregated with designed  $\beta$ -amyloid proteins (Olzscha et al., 2011). Targeting of mitochondria in amyloidosis has been described for  $\alpha$ -synuclein in Parkinson's disease (Haelterman et al., 2014), A $\beta$ (1-40/42) and Tau in Alzheimer's disease (García-Escudero et al., 2013), SOD1 in amyotrophic lateral sclerosis (Taylor et al., 2016), and huntingtin in Huntington's disease (Costa and Scorrano, 2012). A 'mitochondrial side' in amyloid proteinopathies has thus emerged. Overall, in the mitochondria of cells undergoing amyloidosis it is clear that malfunction of the electron transport chain, with subsequent generation of reactive oxygen species (ROS), and the impairment of proton-motive force (PMF), leading to a reduction in the efficiency of ATP synthesis, are major determinants of neurodegeneration (Lin and Beal, 2006; Liu et al., 2015). Since mitochondria have bacterial endosymbiotic ancestry (Gray, 2012), it makes sense to explore if these routes for amyloid toxicity can be reconstructed and untangled in bacteria.

While much information on amyloid diseases is being derived from model systems such as mice, flies, worms, and yeast, which share genetic similarities with humans (Narayan et al., 2014), bacterial cells have not been exploited so much because, when expressed in bacteria, proteins involved in human amyloidosis aggregate as inclusion bodies (IBs) that are barely detrimental to cell fitness (Lindner et al., 2008; Winkler et al., 2010). On the other hand, bacteria use amyloids as functional tools in an extracellular context, e.g., to scaffold biofilms, as in the case of CsgA/*curli* in *Escherichia coli* (Chapman et al., 2002) or TasA in *Bacillus subtilis* (Romero et al., 2010); or to coat aerial hyphae, as chaplins/rodmins in *Streptomyces coelicolor* (Capstick et al., 2011). In particular, the complex secretion pathway for CsgA (Van Gerven et al., 2015) has been exploited as a screening platform to survey the amyloidogenic potential of proteins and to search for inhibitors of amyloidosis (Sivanathan and Hochschild, 2012). Recently, a transcriptional terminator from *Clostridium botulinum* (CbRho), has been characterized as an intracellular prion-like protein (Pallarés et al., 2016; Yuan and Hochschild, 2017). CbRho is the determinant of an epigenetically transmissible phenotype, structurally and functionally analogous to yeast prions (Liebman and Chernoff, 2012), but not a suitable model system for amyloid diseases.

Along the last 10 years, we have developed a synthetic prionoid, i.e., a cytotoxic but non-infectious prion-like protein

(Aguzzi, 2009), by engineering the N-terminal 'winged-helix' domain (WH1) in RepA, the DNA replication protein of a bacterial plasmid (reviewed in Giraldo et al., 2016). As in the full length RepA when activated to initiate DNA replication (Giraldo et al., 2003), RepA-WH1 undergoes a conformational change *in vitro*, coupled to dissociation of protein dimers into monomers, either on transient binding to plasmid-derived DNA sequences (Giraldo, 2007; Gasset-Rosa et al., 2008) or upon templating by RepA-WH1 aggregates themselves (Fernández-Tresguerres et al., 2010). Such process enables the monomers of the highly amyloidogenic mutant A31V of RepA-WH1 to assemble into fibers composed of intertwined tubular helical protein filaments (Giraldo, 2007; Torreira et al., 2015). RepA-WH1 fibers are of amyloid nature, as indicated by Congo red binding (Giraldo, 2007), and by a net increase in the protein  $\beta$ -sheet contents, according to both circular dichroism (Giraldo, 2007; Torreira et al., 2015) and surface-enhanced Raman (Fernández et al., 2016a) spectroscopies. In our efforts to engineer a synthetic bacterial amyloid proteinopathy, we found that the amyloidogenicity of WH1(A31V) in *E. coli* cells can be boosted displacing its conformational equilibrium toward partial unfolding by fusing a protein to its C-terminus, distinct to the natural WH2 domain in RepA (Giraldo et al., 2003): the monomeric fluorescent protein mCherry (Fernández-Tresguerres et al., 2010; Gasset-Rosa et al., 2014; Molina-García and Giraldo, 2014). In the resulting fusion protein, for simplification hereafter WH1(A31V)-mCh (biophysically characterized in Fernández et al., 2016b), the mCherry tag has not a direct contribution to aggregation, because a fusion of mCherry to wild-type RepA-WH1 remained soluble and non-toxic in the cytoplasm (Fernández-Tresguerres et al., 2010; Molina-García and Giraldo, 2014). WH1(A31V)-mCh aggregates are vertically inheritable (from mother to daughter cells) cytotoxic particles (Fernández-Tresguerres et al., 2010), phenotypically distinct to IBs in terms of morphology, intracellular distribution and numbers, higher affinity for an amyloid-specific fluorophore, poor co-localization with IbpA (an IBs-tracer protein), and their acute cytotoxicity (Gasset-Rosa et al., 2014). WH1(A31V)-mCh propagates as at least two amyloid strains (or variants) with distinct morphologies and degrees of cytotoxicity whose interconversion is modulated by the Hsp70 chaperone DnaK (Gasset-Rosa et al., 2014), resembling the phase transitions observed in proteins involved in human amyloidosis (Giraldo et al., 2016). In coherence with the ability of DNA to promote RepA-WH1 amyloidosis *in vitro*, in *E. coli* cells amyloid precursors assemble at the bacterial nucleoid (Moreno-del Álamo et al., 2015). Interestingly, a recent study reveals that the full length RepA protein, through its WH1 domain, assembles as a functional amyloid at the bacterial nucleoid to physically couple plasmid DNA replication origins, thus preventing premature re-initiation events (Molina-García et al., 2016). Binding of WH1(A31V)-mCh to the bacterial cell membrane *in vitro*, or to lipid vesicles having an acidic phospholipid composition, has revealed that lipids also promote the amyloidogenesis of the protein and its assembly as transmembrane pores *in vitro* (Fernández et al., 2016b), as many proteins involved in human amyloidosis do (Butterfield and Lashuel, 2010).

Here we have explored the pathways for the amyloid cytotoxicity triggered by the RepA-WH1 prionoid in *E. coli*, aiming to outline a simplified chain of events shedding light on the molecular mechanism(s) operating in human amyloidoses, which so far have revealed as extremely complex and refractory to untangle. In bacteria undergoing WH1(A31V)-mCh amyloidosis, membrane targeting is operational as the primary mechanism of damage to cells both under aerobic and anaerobic conditions. Combined transcriptomic and interactomic studies reveal that up to 501 genes or proteins are potentially involved in amyloidosis, forming part of over 40 functional clusters of which a significant fraction contributes to the following major cellular processes: carbon metabolism, NADH and (Fe-S)-dependent oxido-reduction, transport through the inner membrane, iron uptake, (Fe-S) clusters assembly, nucleic acids metabolism, cell division and responses to stress, in particular detoxification of ROS. Several of these targets were then functionally validated. The primary loss in PMF leads to a substantial depletion of the ATP pool and, due to the consequent reduction in the intracellular levels of iron, enhances the expression of NdhII. This dehydrogenase generates  $H_2O_2$  by auto-oxidation, while several of the proteins involved in detoxifying peroxide reduce their expression or co-aggregate with the prionoid, thus sensitizing bacteria toward oxidative stress, which ultimately stalls cell division and leads to cell death. RepA-WH1 amyloidosis provides a unique window to survey the essential landscape of a general amyloid proteinopathy, endorsing this prion-like protein as a generic, minimal bacterial model of amyloid disease.

## MATERIALS AND METHODS

### Bacterial Strains and Culture Conditions

Expression of either WH1(A31V)-mCh or WH1( $\Delta$ N37)-mCh was performed from low copy-number plasmids under the control of the *P<sub>tac</sub>* promoter (described in Gasset-Rosa et al., 2014). A construct just carrying the mCherry protein (Molina-García and Giraldo, 2014) was used as a control. As bacterial host, the reduced genome *E. coli* K-12 strain MDS42 *recA* (Pósfai et al., 2006) was used in all experiments because it provides a simplified 'chassis' carrying the essential metabolic and regulatory pathways. Bacterial cells were transformed with the plasmids and grown at 37°C in 200 mL of rich LB medium (supplied with 2 mg·mL<sup>-1</sup> thymine and 100 µg·mL<sup>-1</sup> ampicillin) with good aeration in 1 L Erlenmeyer flasks. Induction was achieved by adding IPTG to 0.5 mM when cultures reached OD<sub>600 nm</sub> = 0.2. Cells were harvested at various post-induction intervals, washed and, for the transcriptomic and interactomic analyses, immediately frozen in liquid nitrogen and then transferred to -70°C for storage. Cells (4·10<sup>8</sup>-3·10<sup>9</sup>, depending on the assay) were collected from at least three independent culture replicates.

### Microscopy

Bacterial cells were observed with a Nikon Eclipse 90i microscope, equipped with a CFI PLAN APO VC 100x (NA 1.40) oil immersion objective and a Hamamatsu ORCA-R<sup>2</sup> CCD

camera. For mCherry fluorescence, a 543/22 nm excitation and 593/40 nm emission filter and 200 ms exposures were used. Differential interference contrast (DIC) shots (100 ms) were also captured. Images were analyzed using the NIS-Elements AR software (Nikon). Bacterial culture aliquots were fixed in formaldehyde and mounted on poly-L-lysine coated slides, as described in Fernández-Tresguerres et al. (2010).

### Luciferase Assays Monitoring Intracellular ATP Levels

In a first approach, *E. coli* bulk cultures, expressing or not the RepA-WH1 prionoid, were grown as indicated above. Upon IPTG induction, every 30 min 4·10<sup>8</sup> bacterial cells were harvested and lysed. The levels of ATP were determined *in vitro* using the ATP Bioluminescence assay HSII (Roche), which is based on the requirement of ATP by firefly luciferase to process luciferin and emit light at 562 nm. Samples were dispensed in 96 wells black-walled microtiter plates and read-outs acquired in a TD-20/20 Turner Designs luminometer. Plots were corrected to the dry weight of cells.

In a second approach, bioluminescence was monitored in real time in microscale cultures. In this assay, bacteria carried the vector for the expression of WH1(A31V)-mCh (Gasset-Rosa et al., 2014) plus mini-CTX-*lux* (Becher and Schweizer, 2000), a plasmid constitutively expressing the *Photobacterium luminescens luxCDABE* operon from the kanamycin promoter. Cultures in LB (no antibiotics added) at OD<sub>600 nm</sub> = 0.05 were fractioned in 200 µL aliquots and displayed in 96 well, flat bottom and black-walled, Grenier Chimney plates. When required, IPTG was supplied to 0.5 mM at the beginning of the experiment and each plate was then incubated in a Tecan infinite M200 PRO plate reader for 24 h at 37°C. At 30 min intervals, plate was agitated for 5 s (2 mm amplitude) and the following variables were sequentially measured: absorption (at 600 nm, 9 nm bandwidth), luminiscence (1 s integration time) and fluorescence (546 nm excitation, 9 nm bandwidth; 600 nm emission, 20 nm bandwidth; 25 flashes for 20 µs). Data were normalized to the OD<sub>600 nm</sub> values. For each experiment, three replicates were set up.

### Determination of the Intracellular Concentration of Iron

Bacterial cultures were grown as specified above and iron concentration in the cell pellets was determined based in the ability of ferrozine to form a complex with Fe<sup>2+</sup> that absorbs light at 562 nm (Honn et al., 2012). Volumes proportional to the cell densities in the cultures (1.0 OD<sub>600 nm</sub> ≈ 8·10<sup>8</sup> bacteria) were taken at time intervals and then cells were harvested, washed and resuspended in PBS buffer. Bacteria were lysed with 100 µL NaOH and then neutralized with 100 µL of 10 mM HCl. Cell lysates were incubated with 100 µL of protein uncoupling solution (0.7 M HCl, 2.25% KMnO<sub>4</sub>) for 2 h at 60°C. Then samples were incubated for 30 min with 100 µL of 6.5 mM ferrozine, 6.5 mM neocuproine, 2.5 M ammonium acetate, 1 M ascorbic acid, and the mixture was centrifuged for 30 s at 13,000 rpm. A<sub>562 nm</sub> was measured for all supernatants in a Varioskan Flash (Thermo scientific) plate reader. The values of



absorption obtained were normalized to the dry cell weight. The whole set of samples was processed at the same time for each replica of the assay to achieve reproducibility.

## Viability of Bacteria Expressing the Prionoid under Aerobic vs. Anaerobic Conditions

Cells were grown aerobically, as described above, or anaerobically in LB medium supplemented with 10 mM nitrate as terminal electron acceptor and 5 mM cysteine as reducing agent. Bottles with 20 ml of LB medium, as well as the nitrate and cysteine stock solutions (100x), were flushed with N<sub>2</sub>, sealed with rubber stoppers and aluminum foil and then autoclaved. Then bottles were introduced in an anaerobic chamber (Forma anaerobic system 1029 S/N, Thermo Scientific) in which the air was continuously interchanged with a mixture of N<sub>2</sub> and biogas (10% H<sub>2</sub>, 5% CO<sub>2</sub> and 85% N<sub>2</sub>). The nitrate and cysteine supplements and the bacterial inocula were injected into the bottles through the stopper and cultures were incubated at 37°C under low shaking conditions (150 rpm). Bacterial growth was monitored as OD<sub>600 nm</sub>. Serial dilutions of the cultures at initial-log phase were plated on LB-agar, which had been supplemented with nitrate and cysteine and left to stand at the anaerobic chamber for at least 24 h before usage. The rest of bacteria were induced with 0.5 mM IPTG and further grown until reaching mid-log and then early stationary phase, when serial dilutions were also plated. Incubations were carried out at 37°C under aerobic or anaerobic conditions and then colony forming units (cfu) per mL were counted. These experiments were performed in triplicate.

## Transcriptomic Analysis of the Response of *E. coli* to the RepA-WH1 Prionoid

WH1(A31V/ΔN37)-mCh expression was induced under aerobiosis as indicated above. For RNA purification, the RNeasy kit (Qiagen) was used, followed by in-column DNaseI digestion (RNase-free, Roche; 10 μL, 2 h at 37°C). The purity of the RNA preparation was assessed first through AGE (0.8% agarose in TAE buffer, samples pre-incubated in 50% formamide buffer, at 95°C for 2 min) and then in a Bioanalyzer 2100 RNA chip (Agilent). Final RNA concentrations ranged between 0.5 and 0.75 μg·mL<sup>-1</sup> and their absorption ratios at 260/280 nm were between 2.13 and 2.45. Equal amounts of each RNA sample were retro-transcribed to DNA using random sequence oligonucleotide hexamers as primers. Template RNAs were then degraded with NaOH and cDNAs were labeled using TdT DNA polymerase and ddUTP-biotin. Labeled cDNAs were hybridized on GeneChip® *E. coli* Genome 2.0 arrays (Affymetrix), which span 10,000 probesets from the pangenome of four *E. coli* strains (including MG1655, the parental for MDS42) and casted on a Fluidics Station 450 (Affymetrix) at 45°C for 16 h. Arrays were washed, stained with phycoerythrin-conjugated streptavidin and then fluorescence emission at 570 nm was digitized in a GeneChip® Scanner 3000 7G (Affymetrix), as specified by the supplier. Microarrays were identically processed for three independent biological replicas. Data were normalized with the RMA algorithm (Affymetrix Expression Console software) and analyzed using the Babelomics

software package (Medina et al., 2010). Statistical analysis of the results was performed through the limma *t*-test with Benjamini–Hochberg's FDR correction: genes with false discovery rates (FDR) ≤ 0.05 were classified as significantly induced/repressed. Data were manually filtered to discard low score (background) genes not present in the MDS42 genome (Pósfai et al., 2006). Genes with A31V/ΔN37 expression ratios either higher than 2 or lower than 0.5 were selected as the fraction of the *E. coli* genome preferentially expressed or repressed, respectively, in response to WH1(A31V)-mCh amyloidosis. Microarray data are available at the Gene Expression Omnibus database (GEO) under the accession number GSE69517.

## Interactomic Analysis of the Co-aggregation of *E. coli* Proteome with RepA-WH1

After induction of MDS42 cells carrying either WH1(A31V)-mCh or WH1(ΔN37)-mCh (see above), 13 A<sub>600 nm</sub> units were processed at 0.5, 1, and 2.5 h by lysing the cell pellets with 1.5 mL of 20 mM Hepes-NaOH pH 6.0, 0.1 M NaCl, 0.5% sulfobetaine 12 (SB-12), 0.5% Na-deoxycholate, 1 mM EDTA, 50 μg·mL<sup>-1</sup> RNaseA, plus a protease inhibitors pill (Roche). Cell lysates were then centrifuged at 12,000 rpm for 1 h at 4°C. Pellets were resuspended in 1.5 mL of the same buffer, but with 1.0 M NaCl and no RNaseA, and they were sonicated (Branson ultrasonic homogenizer, thin tip) for 30 s and centrifuged as above. The sedimented fraction was resuspended in 250 μL of 20 mM Hepes-NaOH pH 6.0, 0.1 M NaCl, 1 mM EDTA and this suspension was then carefully layered on a discontinuous sucrose (20–40% in the same buffer) cushion and centrifuged overnight at 12,000 rpm and 4°C. Pellets were subsequently resuspended in Laemmli buffer (x2), their component proteins analyzed by SDS-PAGE (10% polyacrylamide) and then gels stained with Coomassie blue. Proteins bands over and below WH1(A31V/ΔN37)-mCh were excised, cut into pieces and digested in gel (50 mM NH<sub>4</sub>HCO<sub>3</sub>, overnight at 30°C) with bovine trypsin (12.5 μg·mL<sup>-1</sup>). Peptides were extracted in acetonitrile and 0.5% trifluoroacetic acid, cleaned through a ZipTip (C18 matrix; Millipore) and resuspended in 0.1% formic acid, 2% acetonitrile (buffer-A). Peptides were processed as described (Barderas et al., 2013). Briefly, peptides were trapped in a C18-A1 ASY-Column (Thermo Scientific) and, upon elution, loaded into a Biosphere C18 column (NanoSeparations). A 125-min gradient (250 nL·min<sup>-1</sup>) from 0 to 35% buffer-B (0.1% formic acid in 100% acetonitrile), followed by steps to 45% (15 min) and 95% (10 min), was developed in a NanoEasy HPLC coupled to a nanoelectrospray ion source (Proxeon). Mass spectra (*m/z* 300–1700) were generated in an LTQ-Orbitrap Velos MS (Thermo Scientific) in the positive ion mode and acquired with a target value of 1,000,000 at a resolution of 30,000 (*m/z* 400). The 15 most intense ions were selected for collision-induced fragmentation in the linear ion trap with a target value of 10,000 and normalized collision energy of 38%. Raw MS files were searched with the SEQUEST algorithm (Eng et al., 1994) against the *E. coli* MDS42 proteome (UniProt). Peptides were validated with Percolator (Spivak et al., 2009), scoring as



positive those proteins with  $\geq 3$  identified peptides per target, or with a peptide spectrum match (PSM) value  $\geq$  number of identified peptides and XCorr  $> 3$ . Proteins represented in the mass spectra by a single peptide were not considered, except when PSM  $> 3$ . If present in both datasets, proteins classified as co-aggregated with  $\Delta N37$  were then subtracted from those listed for A31V. The whole procedure was repeated for three independent biological replicas. Proteins found at least twice as preferentially co-aggregated with the A31V variant were selected as the fraction of the *E. coli* proteome co-aggregated with WH1(A31V)-mCh.

## Comparison of the Transcriptomic and Interactomic Datasets

The lists of genes preferentially induced/repressed or co-aggregated with WH1(A31V)-mCh, but not with the  $\Delta N37$  variant, were processed in parallel in a similar way, including Boolean algebra analysis with Venny<sup>1</sup>, classifying genes (or proteins) as *early* when present just in the 0.5 h dataset or when found both at 0.5 and 1.0 h, middle when exclusively placed in the 1.0 h dataset, and late when present at 2.5 h alone or both at 1.0 and 2.5 h. Gene ontology (GO) functional classification was performed with the EcoCyc database (Keseler et al., 2013). The curated transcriptomic and interactomic datasets were finally crossed using the STRING 10.0 tool (Szklarczyk et al., 2015) to get a comprehensive set of the functional pathways and protein clusters involved in WH1(A31V)-mCh amyloidosis.

## HPLC Analysis of Metabolic Succinate and Acetate

Bacterial cultures were grown as indicated above. One mL aliquots were collected at post-induction intervals, cells removed by centrifugation at 13,000 rpm for 5 min, and the culture supernatants were processed through 0.2  $\mu\text{m}$  filters and stored at  $-80^\circ\text{C}$ . Samples were analyzed in triplicate, as described in Felpeto-Santero et al. (2015). Twenty microliter samples were injected into an Aminex HPX-87H column (Bio-Rad) coupled to a Gilson HPLC system. Elution was performed at  $0.6 \text{ mL} \cdot \text{min}^{-1}$  in 5 mM  $\text{H}_2\text{SO}_4$ . Identification and quantitation of the acetate and succinate peaks were carried out using 32 Karat (v. 8.0; Beckman-Coulter). Metabolite concentrations were extrapolated from the elution profiles of calibrated solutions of acetate and succinate. Plots were corrected according to the dry weight of bacterial pellets.

## Assay for Inhibition by ROS of Bacterial Growth on Agar

Bacterial cultures were grown to  $\text{OD}_{600 \text{ nm}} = 0.4$  and  $400 \mu\text{L}$  plated on LB agar with  $100 \mu\text{g} \cdot \text{mL}^{-1}$  ampicillin and 0.5 mM IPTG. When indicated, plates were supplemented with ascorbic acid to 1.5 mM to neutralize hydrogen peroxide. Sterile filter paper disks (Whatman, 0.5 mm  $\varnothing$ ) were embedded with 0.001%  $\text{H}_2\text{O}_2$  or 0.0025% (w/v) paraquat (Sigma), and then laid on the plates and cultured at  $37^\circ\text{C}$  overnight. For the  $\Delta ndh$  SLC22 cells (Woodmansee and Imlay, 2002)  $\text{H}_2\text{O}_2$  was tested up to 0.5%.

<sup>1</sup><http://bioinfogp.cnb.csic.es/tools/venny/index.html>

Areas of the inhibition halos were estimated on photographs, subtracting the area of the paper disks.

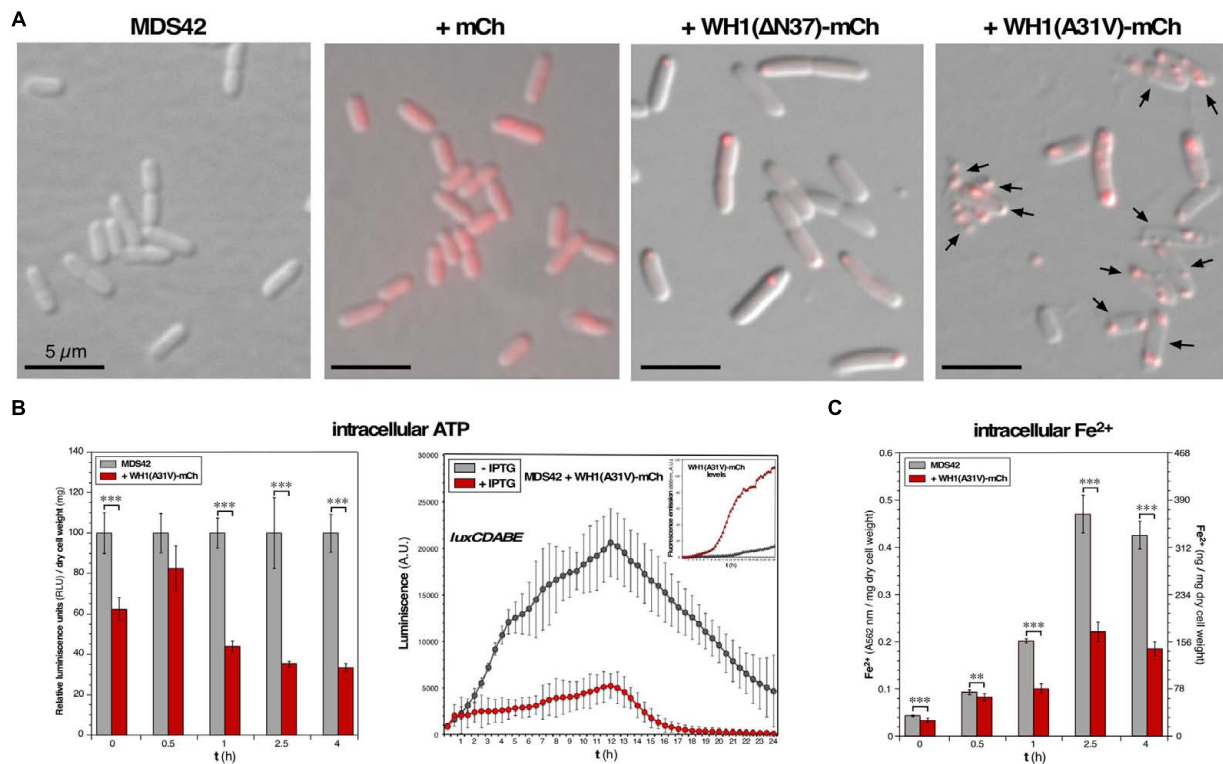
## RESULTS

### WH1(A31V)-mCh Targets the Inner Cell Membrane, Hampering PMF-Dependent Transport and ATP Synthesis

The hyper-amyloidogenic A31V variant of RepA-WH1 (Giraldo, 2007) becomes metastable and highly cytotoxic upon fusion to the monomeric red fluorescent protein mCherry (Gasset-Rosa et al., 2014). The resulting prion-like protein, WH1(A31V)-mCh, has the ability to assemble pores in model lipid vesicles that mimic the *E. coli* inner membrane, thus leaking their contents while not suffering lysis (Fernández et al., 2016b). Expression of WH1(A31V)-mCh in the *E. coli* K-12 MDS42 strain resulted, when bacteria were observed at the microscope (Figure 1A), in a significant proportion of 'ghost' cells. In a clear indication for a weakened integrity of the membrane, cells lost their normal turgor but, as for the vesicles, did not lyse retaining their large size components such as the nucleoid and the prionoid aggregates. On the contrary, bacteria expressing the soluble mCherry reporter did not show any difference in morphology compared with the parental strain.

The integrity of the cell membrane is critical to the generation of a PMF, which drives ATP synthesis by the membrane-bound ATP synthase. If, as observed *in vitro* (Fernández et al., 2016b), WH1(A31V)-mCh targets the inner membrane through pore formation, membrane integrity is expected to be compromised, with the subsequent reduction of PMF-dependent processes such as ATP synthesis. To test this hypothesis, we measured the concentration of ATP in cell lysates from bulk *E. coli* cultures grown aerobically in rich medium, by measuring the *in vitro* activity of the ATP-dependent firefly luciferase: a progressive reduction in luminiscence (up to  $\approx 70\%$  at  $\geq 2.5$  h) was observed upon the expression of the prionoid (Figure 1B, left). In a different bioluminescence assay, based on the constitutive *in vivo* expression of the bacterial *luxCDABE* operon, ATP was consumed by LuxD in the synthesis of the substrate for the LuxAB luciferase. In this case, the expression of WH1(A31V)-mCh also led to a net reduction (by  $\approx 75\%$ ) in luminiscence emission (Figure 1B, right). Both results point to a significant drop in the intracellular amount of ATP and thus are consistent with a scenario of compromised bioenergetics.

We then focused on iron uptake to probe the integrity of the inner cell membrane further. Iron is an essential co-factor in many reactions central to aerobic metabolism, especially those involving the oxidation-reduction of substrates. Being a scarce resource, Gram-negative bacteria have evolved siderophores, scavenger molecules with high-affinity and specificity for iron (Frawley and Fang, 2014). Once synthesized, siderophores are secreted through both membranes and, after extracellular coordination of the  $\text{Fe}^{3+}$  ion, they are internalized in a process



**FIGURE 1 | Functional assessment of membrane integrity in bacteria undergoing the WH1(A31V)-mCh amyloidosis. (A)** Images overlaying red epifluorescence and DIC micrographs of *E. coli* MDS42 cells (left), or the same strain expressing (2.5 h post-induction) mCherry (center), or WH1(A31V)-mCh (right). Arrows: bacterial cells having lost turgor. **(B) Left:** Estimation of the ATP levels in whole cell bacterial lysates. Mean values (bars) and SDs (whiskers) from 3 biological replicas are displayed, normalized to the values measured for prionoid-free cells. **Right:** Effect of the expression of WH1(A31V)-mCh on the luminescence of live *E. coli* cells that constitutively expressed bacterial luciferase. The levels of the prionoid were monitored in parallel by measuring the red fluorescence emission (inset). Dots: mean values from three independent culture wells; whiskers: SDs. **(C)** Intracellular iron levels in naïve bacteria or upon expression of WH1(A31V)-mCh. Iron uptake is impaired in *E. coli* cells undergoing amyloidosis. Y axes represent the read-out of the ferrozine assay (light absorption at 562 nm; left) and its conversion to the amount of ferrous iron (right), both expressed per mg of dry cell mass. Data from 6 biological replicas. One-way ANOVA statistical significance analysis, followed by Tukey's pairwise difference test, was performed for panels B and C. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

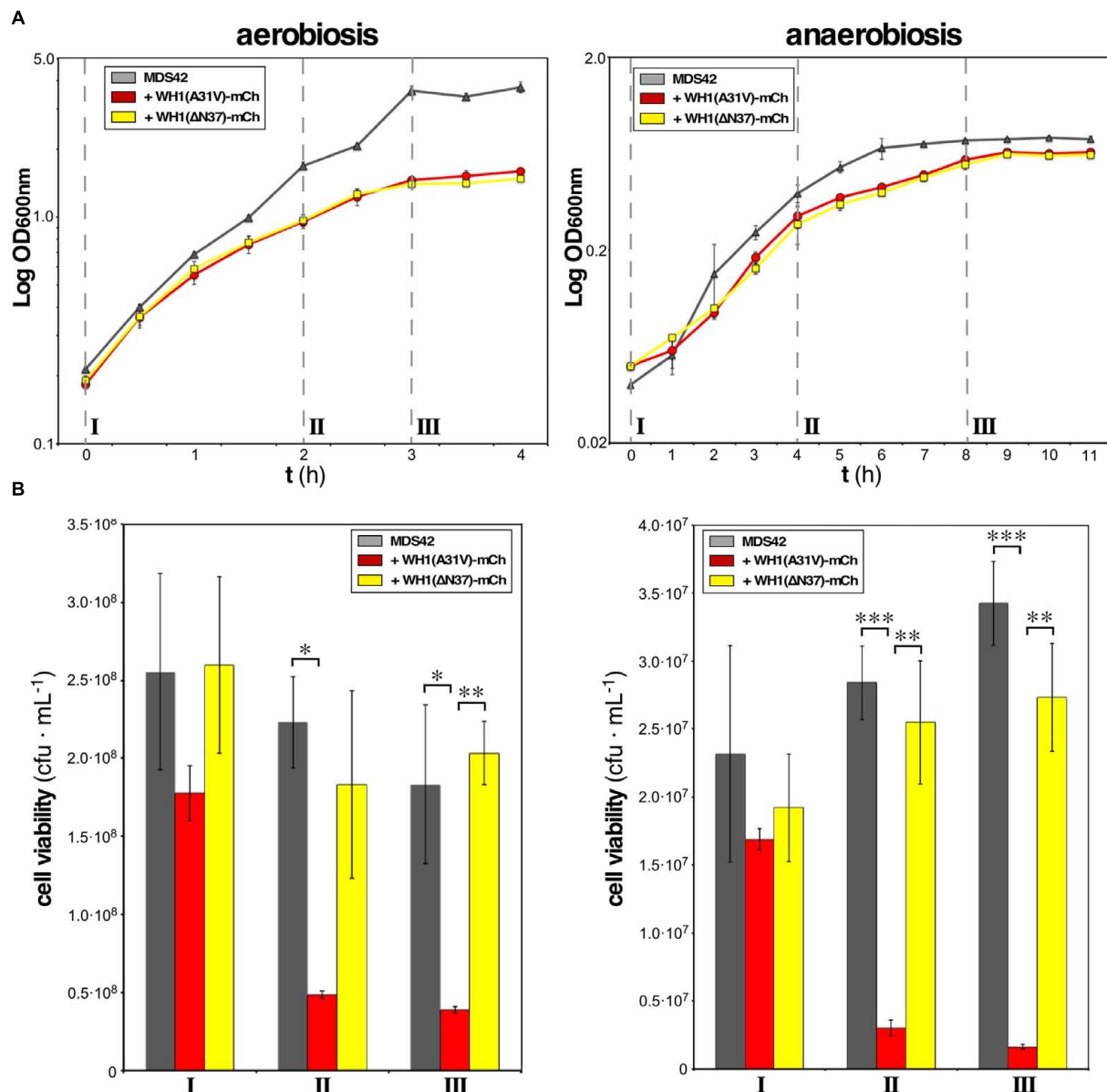
that is dependent on both PMF and ATP consumption. Upon reduction to  $\text{Fe}^{2+}$ , the metal is released in the cytoplasm to be assembled as mononuclear iron or as (Fe-S) clusters in metalloproteins. So, the intracellular level of iron provides another estimation of the ability of the cell membrane to support transport and thus on its integrity. We determined the intracellular concentration of iron across the time course of the induction of WH1(A31V)-mCh (Figure 1C). Ferrous iron increased steadily for 2.5 h in both the naïve and prionoid-expressing cells but was significantly lower (about 50% after 1 h) in the cells undergoing amyloidogenesis.

The findings reported in this section are consistent with a reduction in PMF, and thus in ATP synthesis, due to prionoid-elicited leakage through the inner membrane.

## Viability of *E. coli* Is Reduced by WH1(A31V)-mCh under Both Aerobic and Anaerobic Growth

Targeting of the inner bacterial membrane by amyloids is a mechanism of cytotoxicity that must be operating whatever is

the final acceptor in the electron transport chain. *E. coli* is a facultative anaerobe. Therefore, it made sense to survey whether prionoid cytotoxicity occurred under aerobic and/or anaerobic growth conditions. This study was carried out in parallel upon the expression of two distinct mutant variants of RepA-WH1, A31V and  $\Delta\text{N37}$ , both fused to mCherry: while the former is hyper-amyloidogenic and highly cytotoxic (Giraldo, 2007; Gasset-Rosa et al., 2014), the latter, lacking the amyloidogenic stretch in RepA-WH1, aggregates as conventional IBs and has a milder cytotoxicity (Gasset-Rosa et al., 2014). Relative to the maximum optical density reached by the cells freed of the prionoid, under aerobic conditions a 60% reduction was observed for the cultures expressing the prionoid (Figure 2A, left), whereas in anaerobic growth such reduction was just 20% (Figure 2A, right). The viability of cells was then checked at three points of the respective growth curves: pre-induction, middle exponential and early stationary phases. As expected from the cell densities achieved (Figure 2A), the number of colonies per mL of culture, once plated on agar, was an order of magnitude higher for bacteria grown under aerobiosis than for those in anaerobiosis (Figure 2B). The most noticeable difference was



**FIGURE 2 | WH1(A31V)-mCh expression is cytotoxic under both aerobiosis and anaerobiosis. (A)** Semi-logarithmic growth curves, in aerobic (left) or anaerobic (right) conditions, of naive *E. coli* cells expressing or not either of the two variants (A31V or  $\Delta$ N37) of WH1-mCh. Cultures were grown in triplicate. **(B)** Viability of bacteria from cultures in (A) was checked by plating on LB-agar equal cell numbers at three stages: pre-induction (I), mid-exponential (II), and early stationary (III) phases. Colony forming units (cfu) per mL were counted after incubation under aerobiosis or anaerobiosis. Bars: mean values; whiskers: SDs. One-way ANOVA statistical significance analysis, followed by Tukey's pairwise difference test, was performed (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ).

that, under both physiological conditions, the expression of WH1(A31V)-mCh drastically reduced (to 10–20%) the viability of the bacterial population, whereas WH1( $\Delta$ N37)-mCh did not in a significant way. These results indicated that the expression of WH1(A31V)-mCh indeed is cytotoxic. However, the  $\Delta$ N37 mutant has no deleterious effect and thus the reduction in growth observed for this variant (Figure 2A) must be a burden on fitness imposed by the formation of IBs. As *E. coli* is usually grown under aerobic conditions, and these actually are closer to the environment for human cells undergoing amyloidosis,

the rest of the experiments reported here were carried out in aerobiosis.

### Global Transcriptional Response of *E. coli* to the Expression of the WH1(A31V)-mCh Prionoid

Transcriptomic analysis provided clues on how bacteria react to the expression of the prionoid downstream of its primary target, the inner cell membrane. In a subtractive gene expression

approach using microarrays, WH1( $\Delta$ N37)-mCh IBs were used as a reference set for WH1(A31V)-mCh, thus suppressing from the output list genes involved in the unspecific cellular response to protein aggregation/IBs, such as molecular chaperones and quality control proteases (Winkler et al., 2010). This focused the study on features specific for the acute cytotoxicity of the prionoid. The same *E. coli* strain used above, MDS42 (Pósfai et al., 2006) was selected again as host bacteria because its reduced genome, devoid of non-essential genes, simplified the transcriptomic analysis. In previous studies (Fernández-Tresguerres et al., 2010; Gasset-Rosa et al., 2014), time-lapsed fluorescence microscopy allowed us to characterize 30 min as the post-induction time interval in which WH1(A31V)-mCh aggregates started to become evident in a substantial fraction of the cells, and 2.5 h as the point where cytotoxicity was notorious in the form of stalled cell division, increased filamentation and subsequent cell death, which became dominant at  $\geq 4$  h. We therefore carried out the analysis at 0.5 and 2.5 h, plus an intermediate sampling point (1 h).

Cells from bacterial cultures expressing either WH1(A31V)-mCh or WH1( $\Delta$ N37)-mCh were harvested at the three indicated post-induction times. Total RNA samples were hybridized with DNA microarrays that probed the complete transcriptome of *E. coli*. Differentially expressed genes from the comparison of the A31V and  $\Delta$ N37 datasets were classified as induced ( $\geq 2$ -fold expression level in A31V vs.  $\Delta$ N37, i.e., A31V/ $\Delta$ N37 ratio  $\geq 2.0$ ; in red in **Figure 3A**) or repressed ( $\geq 2$ -fold expression level in  $\Delta$ N37 vs. A31V, i.e., A31V/ $\Delta$ N37 ratio  $\leq 0.5$ ; in green in **Figure 3A**) (**Supplementary Dataset S1**). Genes were then grouped (**Figure 3B**) as *early* expressed (130 genes), if the levels of their mRNAs were altered only at 0.5 h, or both at 0.5 and 1.0 h; *middle* (98 genes), if they appeared in the list exclusively at 1 h; and *late* (145 genes), if they were altered after both 1.0 and 2.5 h, or just at 2.5 h. These three classes comprised most of the genes, with just a few being excluded due to their ubiquitous presence or to their simultaneous clustering at the initial and final datasets. Overall, the *E. coli* transcriptome indicated an initially repressive response to the expression of WH1(A31V)-mCh (86.9% genes differentially repressed vs. 11.5% induced, compared to  $\Delta$ N37, in the *early* group class), with a progressive reactivation of the gene expression program (73.5% genes repressed vs. 26.5% induced; *middle*), which finally became dominant (28.3% genes repressed vs. 69.7% induced; *late*).

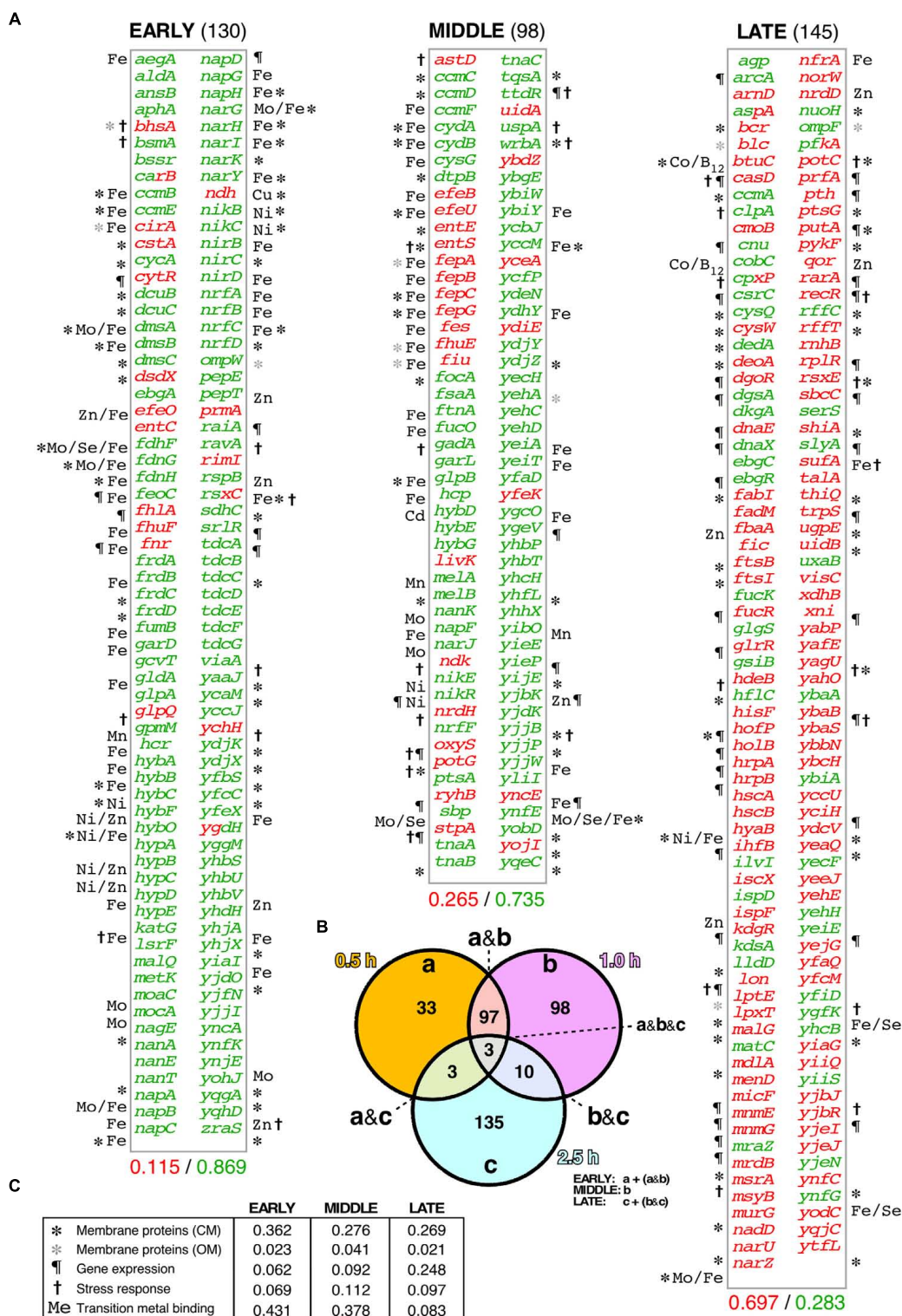
Functional annotation of the genes differentially affected by WH1(A31V)-mCh expression revealed (**Figure 3C**) a major fraction encoding membrane-located proteins, closely followed by metalloproteins, especially at shorter times. Other functional groups included stress response genes and DNA/RNA-binding proteins, such as transcriptional regulators, which became significant in the *late* class accounting for the observed reactivation of gene expression. *Early* repressed genes included many dehydrogenases, terminal reductases and enzymes of the anaerobic metabolism having in common iron as cofactor. This was also the case for the catalase *katG*, a major detoxifier of  $H_2O_2$  (Imlay, 2008, 2013) and the most repressed gene in the whole transcriptomic dataset (**Supplementary Dataset S1**).

Among the few differentially overexpressed *early* genes, were notable those for the synthesis and transport of siderophores (iron uptake pathway), such as *cirA*, *efeO*, *entC*, *fhlA* and *fhuF* (Frawley and Fang, 2014). This response was in agreement with the observed reduction in the levels of intracellular iron (**Figure 1C**). The inductions of the  $H_2O_2$ -responsive gene *ychH* (Lee et al., 2009) and *ndh* were also significant. The latter encodes NdhII, the major NADH-dehydrogenase in exponentially growing *E. coli* cells (Messner and Imlay, 1999), which is typically induced in response to limiting concentrations of intracellular iron (Folsom et al., 2014). On the contrary, other dehydrogenases effective in generating a PMF (Uندن and Bongaerts, 1997) were repressed. The highest *early* expression was achieved for *fnr*, which encodes the oxygen-labile Fe-dependent transcription factor regulating the switch between aerobic and anaerobic metabolism (Myers et al., 2013). The *middle* group class also showed the increased expression of iron uptake genes (*efeBU*, *entES*, *fepABCG*, *fes*, *fhuE*, *fhu*) and of RyhB, an antisense RNA that is the main repressor of genes encoding iron-metalloenzymes (Massé et al., 2005), whereas the gene encoding ferritin (*ftnA*), a major Fe-storage protein, was repressed. In addition, the expression of genes responsible for the response against oxidative stress was enhanced through the regulatory antisense RNA OxyS. Expression of genes for (deoxy)ribonucleotide triphosphate synthesis, such as *ndk* and *nrdH*, and importers of anti-oxidant polyamines like *potG*, was enhanced at the transition to the *late* group class, when cell viability was already severely compromised. Other functional *late* processes included the assembly of (Fe-S) clusters (*hscAB*, *iscX*, *sufA*; being the latter the second highest expressed gene), the responses to osmotic (*putA*) and acidic (*hdeB*) stresses, and elements of the genome maintenance (*deoA*, *holB*, *recR*, *rarA*) and cell division (*ftsBI*, *mrdB*, *murG*) machineries. Relevant to the latter response, it has been recently found that filamented *E. coli* cells with compromised membrane integrity overexpress cell division genes (Sánchez-Gorostiaga et al., 2016).

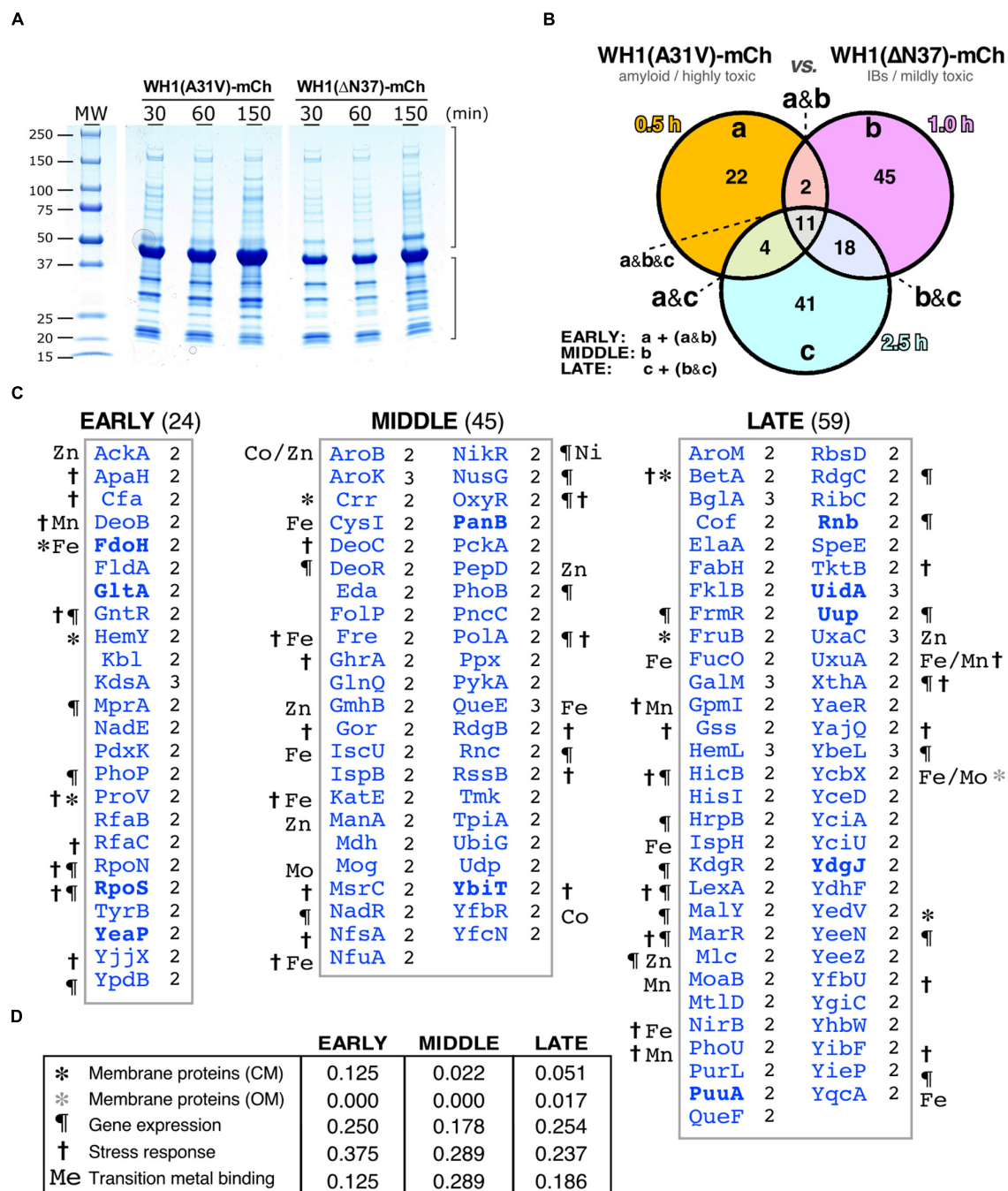
## Assessment of the Fraction of the *E. coli* Proteome Co-aggregated with the WH1(A31V)-mCh Prionoid

The loss of function caused by the assembly of proteins into amyloids is usually associated with co-aggregation of a subset of the cell proteome leading, if not to cytotoxicity itself, to the aggravation of the proteinopathic condition (Olzsch et al., 2011; Hosp et al., 2015). To explore whether the amyloidogenesis of RepA-WH1 led to the differential co-aggregation of particular proteins from the *E. coli* proteome, we undertook the purification and characterization (**Figure 4**) of the aggregated protein subset from bacteria expressing either WH1(A31V)-mCh or its milder version WH1( $\Delta$ N37)-mCh, at the same time intervals previously surveyed through genomic approaches (**Figure 3**). Protein aggregates from three independent cultures were first purified by centrifugation through discontinuous gradients of sucrose, and subsequent separation of the sediment by means of SDS-PAGE (**Figure 4A**). Gel tracks were split into slices and then proteins were digested *in situ* with trypsin. The resulting peptides were





**FIGURE 3 | Differential transcriptomic response of *E. coli* cells to the A31V or the ΔN37 mutants of RepA-WH1. (A)** List of the genes found to be at least two-fold induced (red) or repressed (green) in cells bearing WH1(A31V)-mCh at the early (left), middle and late (right) time intervals, as defined in (B), compared with bacteria carrying the ΔN37 variant (Supplementary Dataset S1). Symbols correspond to the GO terms, as described in (C). The decimal fraction of over-expressed vs. over-repressed genes is printed below. (B) Venn diagram showing the temporal distribution and number of genes whose expression levels were consistently found altered upon induction of WH1(A31V)-mCh. (C) Temporal distribution in five main functional gene ontology (GO) terms of the genes preferentially expressed/repressed with WH1(A31V)-mCh.



**FIGURE 4 | Differential interactomics in *E. coli* cells expressing the A31V or the ΔN37 mutants of RepA-WH1. (A)** SDS-PAGE showing the aggregated protein fraction from bacteria expressing either WH1(A31V)-mCh or the WH1(ΔN37)-mCh mutant. **(B)** Venn diagram displaying the temporal distribution of the proteins found exclusively co-aggregated with the A31V variant of the prionoid. **(C)** Lists of the proteins found co-aggregated with WH1(A31V)-mCh, but not with ΔN37, in at least two of the three biological replicates (right-hand notation: 2/3; **Supplementary Dataset S2**) at the indicated time slots **(B)**. Symbols correspond to the GO terms, as described in panel D (to compare with transcriptomics, see **Figure 3C**). The eight proteins in common with the transcriptomic dataset (**Figure 3**) are in boldface. **(D)** Temporal distribution across five main functional gene ontology (GO) terms of the proteins preferentially co-aggregated with WH1(A31V)-mCh.

extracted and analyzed by nano-scale HPLC combined with mass spectrometry (ESI-MS). Peptides were identified in sequence databases and then classified (**Figure 4B**) following the same criteria used for the microarray studies (**Figure 3B**).

Proteins identified as preferentially co-aggregating with the WH1(A31V)-mCh prionoid, but not with the WH1(ΔN37)-mCh IBs, (**Figure 4C** and **Supplementary Dataset S2**) were less than those inferred from the transcriptomic studies (**Figure 3**):

24 proteins were consistently found (i.e., they were present with a significant score in at least two out of three biological replicas) at the *early* time interval of expression, 45 at the *middle* class and 59 at the *late* group (Figure 4C). Overall, functional annotation revealed that membrane proteins were underrepresented in the datasets, as expected for cytoplasmic aggregates, whereas proteins involved in the response to different types of stress were overrepresented, albeit decreasing along the time course, with the gene expression and transition metal binding functional classes ranking second and third, respectively (Figure 4D). The master regulator of the general stress response RpoS ( $\sigma^{38}/\sigma^S$ ) (Battesti et al., 2011) was among the factors aggregating at the *early* time interval. In the *middle* group, the RpoS inhibitor RssB was found together with a number of proteins involved in the response to oxidative stress such as its master regulator OxyR (Aslund et al., 1999; Seo et al., 2015), the alternative catalase KatE, and the glutathione reductase Gor (Imlay, 2008, 2013). The (Fe-S) cluster scaffolding proteins IscU and NfuA (Jang and Imlay, 2010) were also placed in this subset. In the *late* class, BetA, an enzyme for the synthesis of the osmo-protectant betaine (Lamark et al., 1996) and DNA repair enzymes such as RdgC and XthA were identified. It is noteworthy that several enzymes in the glycolytic (pyruvate kinase II, PykA; triosephosphate isomerase, TpiA), TCA (malate dehydrogenase, Mdh) and mixed acid fermentation (acetate kinase, AckA) pathways appeared aggregated with RepA-WH1(A31V)-mCherry at the *early* and *middle* subsets.

## Combining Transcriptomics and Interactomics Highlights Central Pathways in WH1(A31V)-mCh Amyloidosis

The lists of genes up/down regulated in the transcriptomic analysis (Figure 3) and of proteins found as preferentially co-aggregated with WH1(A31V)-mCh (Figure 4) were then compared. The assumption was that differential gene expression and protein sequestration might be independent contributors to RepA-WH1 amyloidosis and thus complementary, rather than overlapping, views to the core cellular processes involved in the ‘disease.’ Indeed only eight proteins, and their respective genes, were present in both ‘omic’ datasets (1.6% of a total of 501).

Network analysis of the combined set of genes or proteins allowed their assignment to over 40 functional clusters (Figure 5), which could be broadly grouped into eight core functions: hydrocarbon metabolism, respiration [i.e., electron transport, NAD(P)H oxidoreductases and hydrogenases]; nucleotide/phosphate and nucleic acids metabolism; transport through membranes; cell division; iron uptake; (Fe-S) clusters biogenesis; and response to various stresses (with a focus on detoxification of hydrogen peroxide). In terms of the regulatory response(s) to the aggregation of WH1(A31V)-mCh, the analysis of the combined transcriptomic and interactomic datasets revealed that the master regulators of the transcriptional switches in response to oxygen levels, Fnr (Myers et al., 2013), and to general stress, RpoS (Battesti et al., 2011), were directly controlling the expression of substantial fractions (16.21 and

6.64%, respectively, with 1.95% regulated by both) of the genes linked to RepA-WH1 amyloidosis (Figure 6). Other transcription factors, such as OxyR, ArcA, Fur, RpoN/E, PhoB, LexA or CpxR, fell well behind.

The assays presented above converge in a picture of damage to the bacterial inner membrane by WH1(A31V)-mCh with the subsequent reduction in PMF-dependent transport of metabolites and co-factors, such as iron. Limiting iron levels would promote the expression of the NdhII dehydrogenase that, under aerobic conditions, would generate ROS, while a battery of the proteins responsive to oxidative stress would become disabled by co-aggregation with the prionoid. With the aim of validating this sketch of the bacterial amyloidosis, we undertook additional functional assays in *E. coli* cultures expressing WH1(A31V)-mCh or WH1( $\Delta$ N37)-mCh under the same conditions surveyed through the genomic and interactomic approaches.

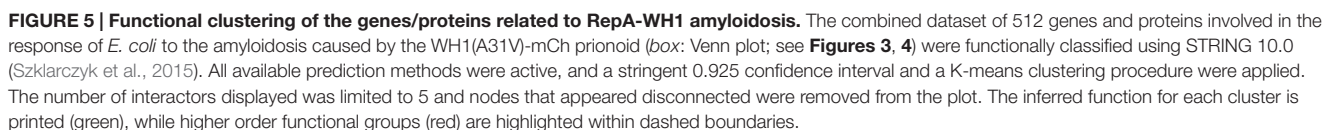
## WH1(A31V)-mCh Amyloidosis Leads to Impaired Carbon Metabolism

Replenishment of ATP from ADP has other sources apart from ATP synthase: the reactions of the central carbon metabolism and substrate-level phosphorylation. Upon the impairment of ATP synthase activity due to the disruption of PMF by membrane leakage, cells would become dependent on less efficient metabolic fluxes (see above; Figure 1B). Thus, glycolysis must be enhanced, as suggested by the observed induction of the pyruvate kinase gene (*pykF*) in cells expressing WH1(A31V)-mCh (Figure 3A). However, this does not seem to be the case probably due to the co-aggregation with the prionoid of triosephosphate isomerase (TpiA; Figure 4C). Therefore, other alternative sources for ATP regeneration were explored.

Determination by HPLC of the extracellular levels of succinate, a key intermediate in the TCA cycle, showed that expression of the prionoid resulted in a net 30% decrease in this metabolite after 1 h (Figure 7A). This probably reflects an early blockage in the TCA cycle due to the co-aggregation of malate dehydrogenase (Mdh) with WH1(A31V)-mCh (Figure 4C), besides the impossibility to regenerate  $\text{NAD}^+$  at the level of the electron transport chain. Interestingly, succinate levels remained more elevated for the  $\Delta$ N37 than for the A31V variant of RepA-WH1, resembling the behavior of wild-type cells.

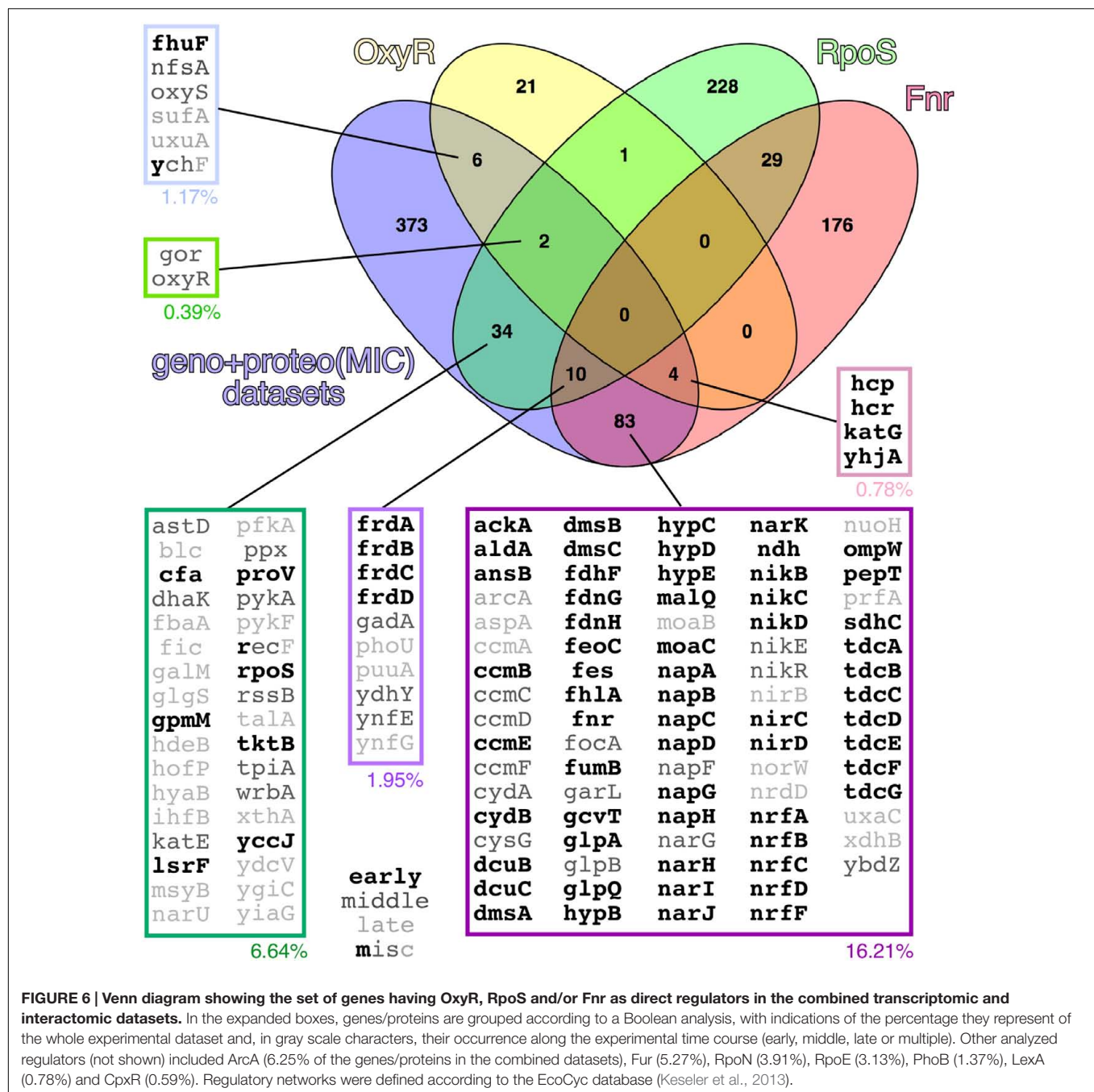
In rapidly growing *E. coli* cells, potentially leading to oxygen-limiting conditions, extra reducing power and ATP are usually generated through mixed-acid fermentation, whose products (acetate in particular) are secreted to the medium but, upon reaching stationary phase, are imported to be further metabolized (Förster and Gescher, 2014). Such double-way metabolic flux was observed in the HPLC determination of the levels of acetate in the culture medium of bacteria not expressing the prionoid (Figure 7B). However, upon the expression of WH1(A31V)-mCh a significant decrease (up to 30%) in the levels of acetate was detected at 2.0–2.5 h post-induction, suggesting a reduction in the production of ATP at substrate-level phosphorylation. This fact could be due to the co-aggregation of acetate kinase (AckA) in the *early* interactomic dataset (Figure 4C), but also





Overall, these results are consistent with a primary disruption in PMF by the RepA-WH1 prionoid, reinforced by a net



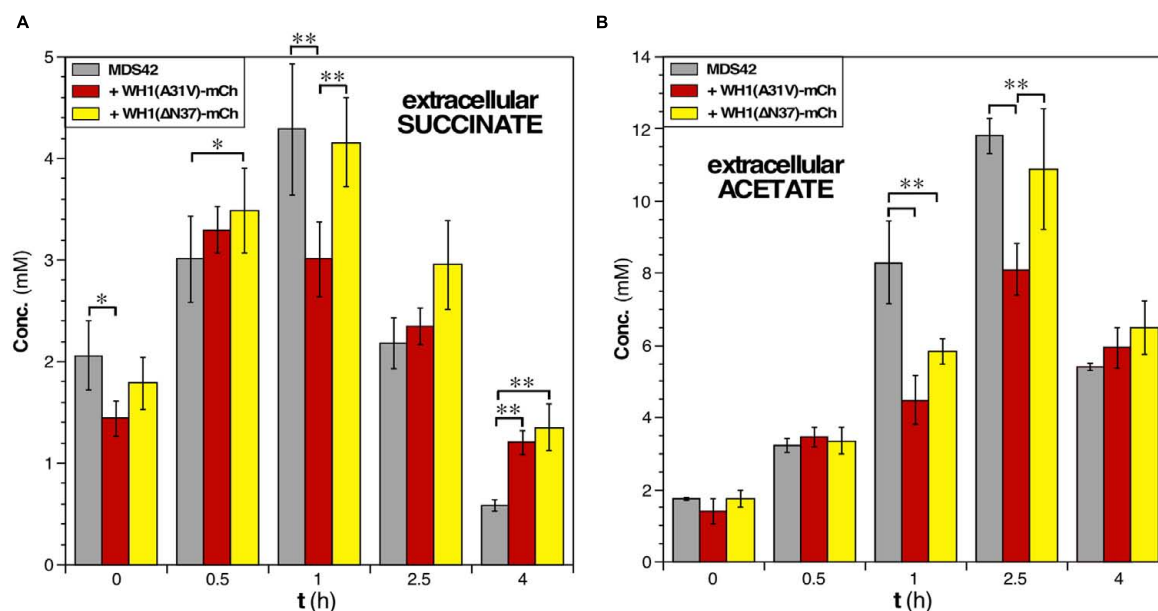


reduction in the fluxes through both central carbon metabolism and mixed acid fermentation.

## WH1(A31V)-mCh Amyloidosis Sensitizes Bacterial Cells to Hydrogen Peroxide

Our transcriptomic analysis on *E. coli* cells in aerobiosis had shown that *katG*, the gene coding the major catalase/peroxidase at the exponential growth phase (Imley, 2008, 2013), was the most repressed at the *early* time interval upon WH1(A31V)-mCh expression (Figure 3A and Supplementary Dataset S1). In

addition, interactomics had identified the alternative stationary phase catalase KatE as significantly trapped in the intracellular aggregates of the prionoid (Figure 4C and Supplementary Dataset S2) (Imley, 2008, 2013). These observations meant that *E. coli* cells suffering from WH1(A31V)-mCh amyloidosis must exhibit increased sensitivity toward stress by hydrogen peroxide. On the contrary, no superoxide dismutase (SodABC) showed altered expression, or differential co-aggregation, upon expression of the A31V or ΔN37 variants. Therefore, bacterial cells undergoing the WH1(A31V)-mCh amyloidosis must not be differentially sensitive to superoxide.



**FIGURE 7 | Assessment of the concentrations of key metabolites in *E. coli* cells undergoing the RepA-WH1 amyloidosis. (A)** Determination of the extracellular concentration of the TCA cycle intermediate succinate in cultures of naïve cells or expressing the two variants of WH1-mCh. Histogram plot the concentrations (areas under HPLC peaks) of the metabolite at the indicated post-induction times, upon correction for the dry cell mass of the bacteria in the cultures. The mean values from 3 biological replicas and the SDs are shown. **(B)** The extracellular concentration of the fermentative end metabolite acetate, determined as indicated in A. One-way ANOVA plus Tukey's pairwise difference test were used to check for statistical significance. \* $p < 0.05$ ; \*\* $p < 0.01$ .

We thus challenged bacteria with diluted  $H_2O_2$  (Figure 8A, left) or paraquat (Figure 8A, middle), a generator of superoxide radicals ( $O_2^{\bullet-}$ ), and tested their effects in a zonal growth inhibition assay on agar plates. Briefly, lawns of cells expressing the control marker mCherry, or its fusion to WH1(A31V) or WH1(ΔN37), were seeded just before laying filters pre-embedded with the oxidizing agents. Quantitation of the areas of the inhibition halos revealed (Figure 8A, right) that the expression of WH1(A31V)-mCh correlated with a net hindrance of bacterial proliferation by  $H_2O_2$  (125% increase in area, compared with the mCherry control), an inhibition higher than that observed upon expression of the ΔN37 variant (43% increase). However, no significant differences were appreciated when the three bacterial strains were treated with paraquat. As expected, the inhibitory effect of  $H_2O_2$  was relieved by the inclusion of a reducing agent (ascorbate) in the medium (Figure 8A, left).

These results support an impairment, dependent on WH1(A31V)-mCh, of the cellular response against the oxidative stress caused by  $H_2O_2$ .

### NdhII Likely Is a Source of ROS in *E. coli* Cells Undergoing WH1(A31V)-mCh Amyloidosis

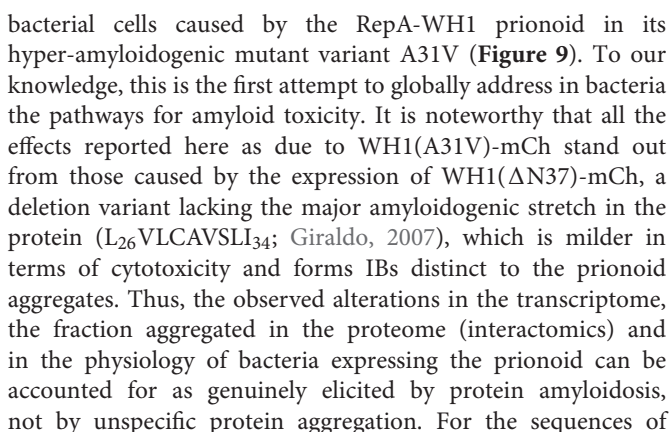
Growth of bacteria under aerobic conditions generates vast amounts of ROS (up to  $\mu M$  intracellular concentrations) (Messner and Imlay, 1999). In cells undergoing WH1(A31V)-mCh amyloidosis, the only differentially induced NADH-dehydrogenase at the *early* stage was NdhII (encoded

by *ndh*; Figure 3A). NdhII is expressed in response to limiting levels of iron (Folsom et al., 2014) and generates ROS through the auto-oxidation of its FAD cofactor (Messner and Imlay, 1999; Woodmansee and Imlay, 2002; but see Seaver and Imlay, 2004). So, an increase in oxidative stress was expected as a side effect of the observed rise in the intracellular levels of NdhII, with the possible consequence of a higher sensitization of cells to exogenous oxidizing agents. To test this hypothesis, zonal growth inhibition assays with  $H_2O_2$  were performed in a  $\Delta ndh$  (null) mutant *E. coli* background (Woodmansee and Imlay, 2002). The results (Figure 8B, left) revealed a net reduction in the sensitivity of the mutant bacteria to the additional stress imposed by exogenous hydrogen peroxide: up to a 500-fold increase in  $H_2O_2$  concentration was required to get inhibition halos with an area close to that observed in the  $ndh^+$  background, while keeping the trend of the higher sensitivity of bacteria expressing WH1(A31V)-mCh (Figure 8B, right).

These results suggest that induction of the alternative dehydrogenase NdhII is a relevant source of ROS in bacteria undergoing WH1(A31V)-mCh amyloidosis, to the point of overtaking proteins involved in detoxifying  $H_2O_2$ , a defense line already feeble due to their co-aggregation with the prionoid (Figure 4C).

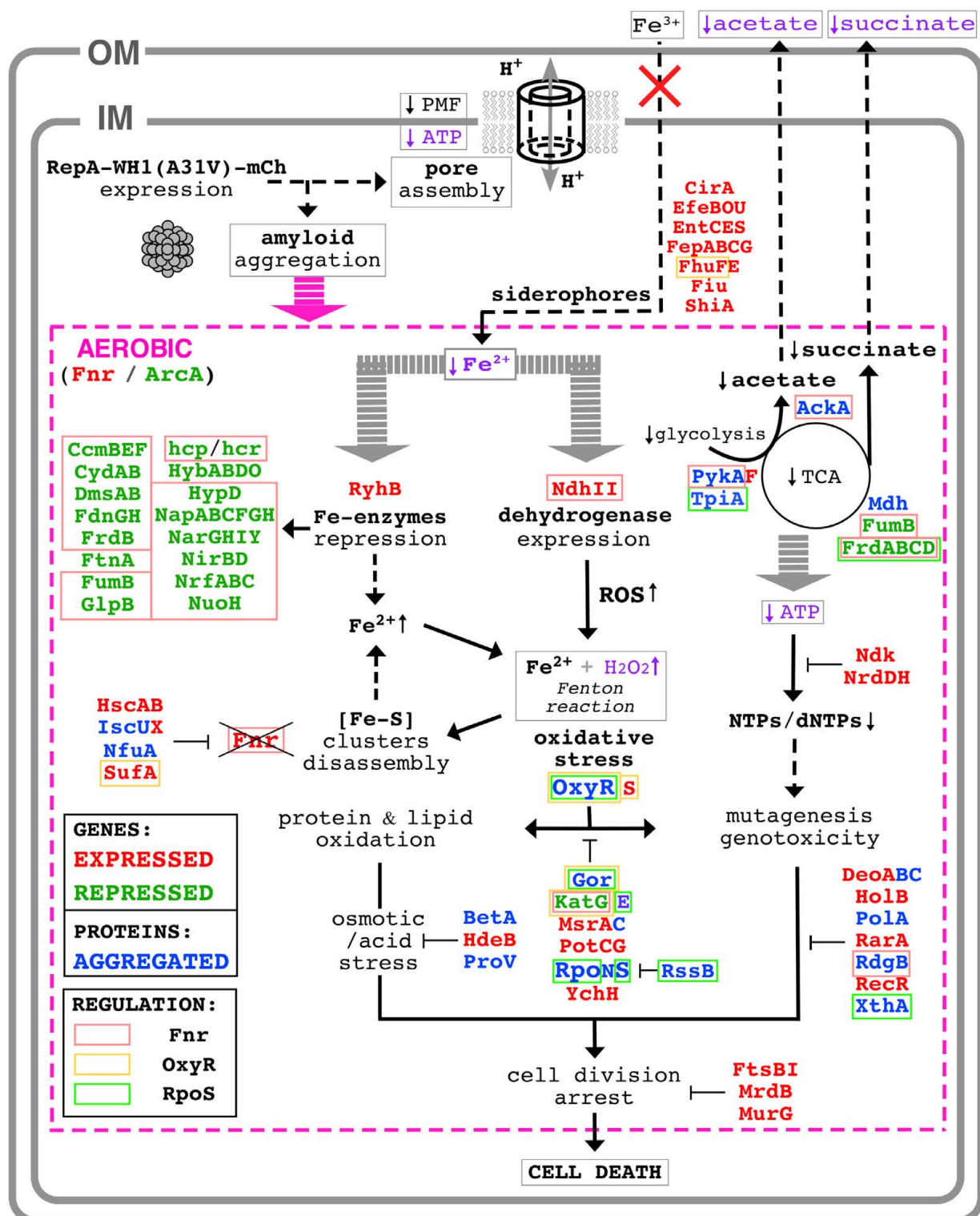
## DISCUSSION

Through a combination of complementary approaches, we have outlined a chain of events leading to the death of



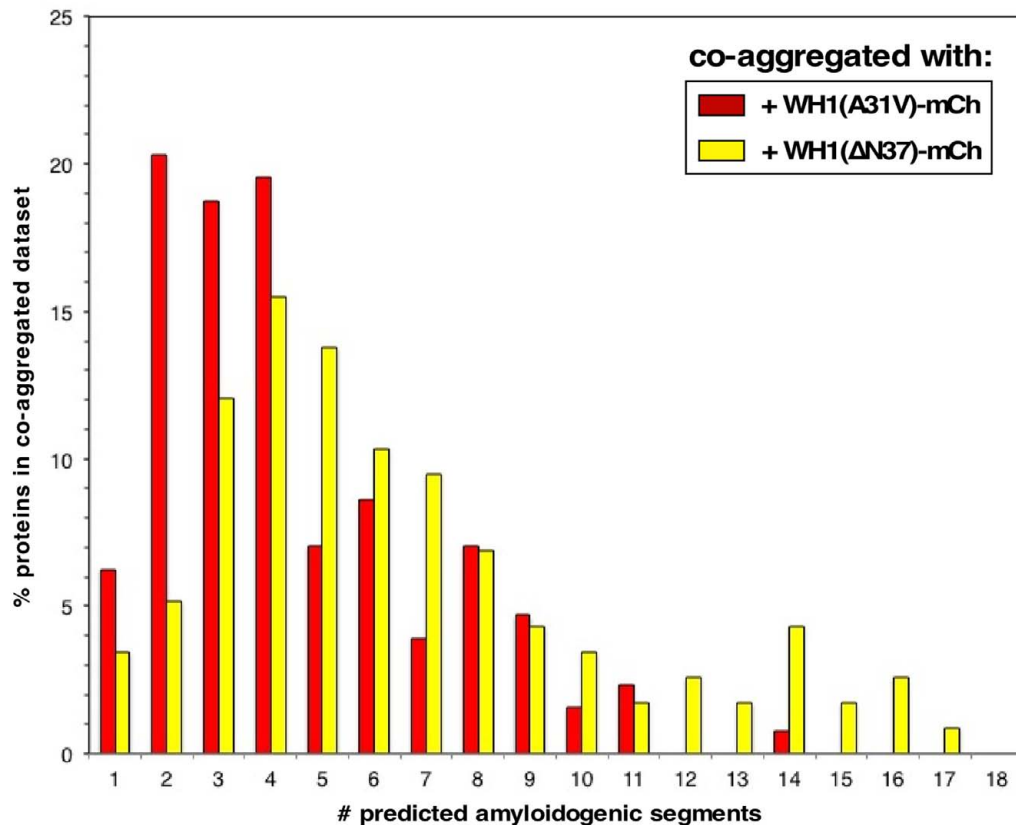
the proteins differentially co-aggregated with WH1(A31V)-mCh, the distribution of predicted aggregation-prone stretches clusters around 2–4 per protein, while those aggregated with WH1( $\Delta$ N37)-mCh show a more spread, bimodal distribution around 4–5 and 14 stretches (**Figure 10**). A similar trend had been described while comparing the sequences of proteins involved in amyloid diseases with those aggregating as IBs, and it was attributed to the ability of amyloids to assemble on the basis of a defined and discrete number of interfaces, instead of the multiple, barely specific contacts established in IBs (Conchillo-Solé et al., 2007). The amyloidogenic stretch in WH1(A31V) might capture, while assembling, other amyloidogenic segments in the proteome, whereas WH1( $\Delta$ N37) would entrap less selectively other proteins, through multiple





**FIGURE 9 | Scheme of the molecular pathways leading to bacterial cell death by the RepA-WH1 prionoid.** The intracellular WH1(A31V)-mCh prionoid drills pores through the bacterial inner membrane, thus triggering a proteinopathy. For further details on the downstream aerobic pathways (magenta), see text. Proteins whose expression was found enhanced (red) or reduced (green) in an attempt to counteract the effects of amyloidosis are indicated, as well as those co-aggregated (blue) with the prionoid. The latter are expected to be functionally defective, worsening the course of the 'disease.' The functional assays performed here to validate pathways picked out by the 'omic' approaches are typed in purple. Three master regulators of the response to stress (Fnr, OxyR and RpoS) appear engaged in WH1(A31V)-mCh amyloidosis, and thus the proteins they regulate are displayed boxed (Figure 6).





**FIGURE 10 |** Distribution of the number of amyloidogenic sequences in the proteins exclusively co-aggregated with WH1(A31V)-mCh (red) or with the ΔN37 mutant (yellow) (Figure 4 and Supplementary Dataset S2). The sequences of the proteins found to co-aggregate with the A31V variant (128 entries) and those found in ΔN37 aggregates (116 proteins) were analyzed with the WALTZ tool to predict amyloidogenic stretches in proteins (Maurer-Stroh et al., 2010).

hydrophobic interactions, while they are folding. It is remarkable that the entries in the transcriptomic and interactomic datasets show little overlap, as expected if co-aggregation with, and transcriptional response to, the prionoid were additive players in RepA-WH1 amyloidosis.

The synthetic bacterial model ‘proteinopathy’ caused in *E. coli* by WH1(A31V)-mCh would be initiated upon targeting the prionoid to the inner cell membrane, in coherence with recent results on the assembly by this protein of pores through lipidic vesicles *in vitro* (Fernández et al., 2016b). The assembly of membrane pores is common to several amyloidogenic proteins involved in human neurodegenerative diseases (Butterfield and Lashuel, 2010). Although leakage through the pores of small molecule cofactors essential for the respiratory chain cannot be excluded, membrane drilling necessarily leads to disruption of PMF. A shutdown of PMF is consistent with the observed reduction in the coupled transport of iron (Figure 1C), and would hinder the activity of transmembrane dehydrogenases (including Complex I: NdhI/NuoA-N). Membrane damage is the primary physical mechanism of toxicity and it is operational in both aerobiosis and anaerobiosis (Figure 2).

In a scenario of low PMF due to a leaky inner membrane, respiration would heavily depend on the alternative NADH

dehydrogenase NdhII, which is induced (Figure 3A) in response to low intracellular levels of iron (Folsom et al., 2014), as found upon WH1(A31V)-mCh expression (Figure 1C). NdhII usually is the most active dehydrogenase under exponential aerobic growth (Messner and Imlay, 1999), but the expression of the prionoid seems to potentiate further such central role. However, NdhII has its disadvantages. Firstly, because it does not create a PMF, NdhII is poorer than NdhI in terms of generation of ATP (Uندن and Bongaerts, 1997). Rather than depending on the normal end of the respiratory chain ( $F_1/F_0$  ATPase), bacteria would then rely on the glycolytic pathway to recharge ATP from ADP. However, the energetic metabolism in bacteria undergoing the WH1(A31V)-mCh amyloidosis is affected by a decrease in the flux from glycolysis to the TCA cycle (Figure 7A), and also in substrate level phosphorylation (i.e., acetate fermentation; Figure 7B). Such restrictions in metabolic fluxes might be imposed by the co-aggregation with the prionoid of key enzymes such as TpiA, PykA, Mdh and AckA (Figure 4C). There is an attempt to regenerate the pool of nucleotide triphosphates through the overexpression of the nucleotide di-phosphate kinase Ndk, but this must be inefficient because this enzyme uses ATP. Secondly, as a by-product of NdhII activity, vast amounts of ROS (both superoxide and hydrogen peroxide) are

generated (Messner and Imlay, 1999; Seaver and Imlay, 2004). NdhII seems to be a relevant source of oxidative stress in cells undergoing WH1(A31V)-mCh amyloidosis, as revealed by the enhanced sensitivity to a challenge with exogenous  $\text{H}_2\text{O}_2$  in *ndh*<sup>+</sup> (Figure 8A) over  $\Delta$ *ndh* genetic backgrounds (Figure 8B). Superoxide dismutases (SodAB) seem to be unaltered in the transcriptome (Figure 3) and are absent from the co-aggregated proteome (Figure 4), thus they cope with the transmutation into  $\text{H}_2\text{O}_2$  of the  $\text{O}_2^{\bullet-}$  radicals generated by NdhII. However, the oxidative stress-responsive catalases are either hyper-repressed (KatG) or co-aggregated with WH1(A31V)-mCh (KatE), thus converting  $\text{H}_2\text{O}_2$  in a major problem. The repression of *katG* can be due to the aggregation of OxyR, which is the master transcriptional activator of the genes responsive to oxidative stress (Aslund et al., 1999; Seo et al., 2015; Imlay, 2015b). This would also explain why other members of the OxyR regulon, such as *ahpCF*, *dps* and *fur*, do not show up in differential transcriptomics (Figure 3A). It is noteworthy that simultaneous disabling of several detoxifying enzymes, as implied here from the sequestering of OxyR, KatE and Gor into the aggregates, has been postulated as a requirement in sensitizing bacteria against ROS (Imlay, 2015a).

Another consequence of an uncontrolled generation of ROS is the  $\text{H}_2\text{O}_2$ -promoted disassembly of (Fe-S) clusters. In particular, the transcriptional regulator Fnr is a sensible target in oxidative stress (Myers et al., 2013). Fnr turnover seems to be assured through an increase in its transcription, as *fnr* is in fact the most expressed *early* gene (Supplementary Dataset S1). In the combined genomic and interactomic datasets, up to 62 *early* genes/proteins (40.79% of 152) are directly regulated by Fnr, whereas this number goes down to 20 (14.49% of 138) and 13 (6.53% of 199) in the *middle* and *late* groups, respectively (Figure 6). Therefore, Fnr likely is the transcription factor responsible for triggering the transcriptomic response of *E. coli* cells to the expression of the WH1(A31V)-mCh prionoid.

In the global transcriptional response to the WH1(A31V)-mCh amyloidosis (Figure 3 and Supplementary Dataset S1) it is noteworthy the induction of genes encoding siderophores, iron scavengers that are first exported and then internalized through the two *E. coli* membranes to fulfill their role (Frawley and Fang, 2014). Since such transport is actually impaired (Figure 1C) due to the reduction in PMF and ATP levels (Figure 1B) imposed by membrane leakage, siderophore expression most likely is futile. Bacteria also seem to react to iron starvation by repressing a plethora of metabolic enzymes having this metal as a cofactor, through the expression of the small antisense RNA RyhB (Massé et al., 2005). A second source that may increase the availability of iron is disassembly of the essential (Fe-S) clusters that, as mentioned above for Fnr, is enhanced by oxidative stress and would be counteracted by induction of proteins involved in chaperoning their assembly, such as IscX, HscAB and SufA (Jang and Imlay, 2010) (Figure 3A). However, this route might be compromised because IscU and NfuA were found co-aggregated with the prionoid (Figure 4C).

In the late stage of the synthetic amyloidosis caused in *E. coli* by the WH1(A31V)-mCh prionoid, the concurrence in the cytoplasm of  $\text{H}_2\text{O}_2$  and some freed iron, the latter from

dismantled mononuclear Fe-enzymes and (Fe-S)-containing proteins and the reduced levels of a major Fe-storage protein (ferritin, FtnA), would result in the generation of hydroxyl radicals through Fenton chemistry. These radicals lead to massive oxidation of lipids, proteins and DNA, and the outcome of genotoxicity (Al Mamun et al., 2012). Although this final sequence of events remains to be experimentally addressed, it seems that there is a last attempt of counteracting such a 'terminal multi-systemic failure' through expression of a battery of enzymes in the response pathways to oxidative, osmotic and acidic stresses, as well as involved in DNA repair and cell division (Figure 3). However, such desperate efforts had no apparent success, since bacteria were committed to death since the initial targeting of the cell membrane.

The sequence of events sketched above for the WH1(A31V)-mCh amyloidosis in *E. coli* (Figure 9) has some points in common with the phenotypic responses that this bacterium assembles to confront, besides oxidative stress (Myers et al., 2013; Seo et al., 2015), other kind of injuries such as acidic pH and osmotic/salt stresses (Weber et al., 2005), high pressures (Malone et al., 2006), iron starvation (Folsom et al., 2014), phage/envelope stress (Bury-Moné et al., 2009), stress-induced mutagenesis (Al Mamun et al., 2012), and antibiotic treatment (Foti et al., 2012). Probably the mechanism closest to that proposed here for the RepA-WH1 prionoid is found for cationic antimicrobial peptides, which target cell membranes as amyloids do and trigger a similar ROS response (Choi et al., 2015). It is noteworthy that some of the routes outlined here for amyloid toxicity, in particular those relative to membrane bioenergetics and central metabolism, have been described as relevant for bacteria to become 'persisters' against external stress, including antibiotics (Harms et al., 2016). However, as a viable state, persistence can be overcome thanks to the stress-responsive genes regulated by RpoS, while in WH1(A31V)-mCh amyloidosis this transcription factor is early sequestered through aggregation (Figure 4C). The cytotoxicity elicited by the bacterial prionoid thus appears to be a class of its own.

Interestingly, the scenario outlined for the bacterial WH1(A31V)-mCh amyloidosis (Figure 9), far from being an oddity emerging from a synthetic construction, might resemble some mitochondrial routes in a wide spectrum of human amyloid diseases (Lin and Beal, 2006). Although mammalian cells lack the alternative NdhII dehydrogenase, A $\beta$ , Tau and  $\alpha$ -synuclein induce the generation of ROS by Complex I (NdhI) in neurons and glial cells, with the impairment of transport through membranes and a reduction in ATP generation (Liu et al., 2015). WH1(A31V)-mCh amyloidosis also shares significant similarities with the cytotoxicity pathways described for PrP in transmissible spongiform encephalopathies: (i) the generation of ROS in glial cells by NAD(P)H oxidase (NOX2) in the respiratory chain (Sorce et al., 2014); and (ii) the expression of genes involved in iron homeostasis (Hwang et al., 2009).

The data presented here on the molecular pathways of the 'proteinopathy' caused in bacteria by the prionoid WH1(A31V)-mCh outline a minimal, reductionist sketch for

a general amyloid disease at the cellular level that, as main core dysfunctions, would imply: (i) protein aggregates targeting the bacterial (or mitochondrial) inner membrane, linked to impaired transport and respiration; and (ii) the subsequent iron-enhanced generation of cytotoxic ROS, coupled to co-aggregation driven inactivation of key detoxifying proteins. Adding to the discoveries made along the last decade on this prion-like protein, the results reported here empower bacteria as model systems of amyloidosis, providing a versatile platform to test interventions aiming to counteract intracellular amyloid proteinopathies in more complex systems.

## AUTHOR CONTRIBUTIONS

LM-G, MM-d, PB, ZM-M, MF, AA-d and AS-G performed the research; JG-C and JN designed the transcriptomic and metabolite analyses, respectively; all authors analyzed data; RG conceived the project, integrated the results and wrote the paper.

## FUNDING

This work has been supported by grants from Spanish AEI/EU-FEDER (BIO2012-30852, BIO2015-68730-R and CSD2009-00088) and CSIC (i-LINK0889) to RG. The publication fee has been paid in part by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

## ACKNOWLEDGMENTS

We thank the members of the Synthetic Microbial Macromolecular Assemblies group at CIB-CSIC for much

encouragement. We are indebted to Jim Imlay (University of Illinois, USA) for providing us with the  $\Delta ndh$  strain and for inspiring suggestions on oxidative stress, and Eduardo Rial and Eduardo Díaz for valuable discussions on bioenergetics and anaerobic growth of bacteria, respectively. The help of Carmen Felpeto and Olga M. Revelles with HPLC is also acknowledged. The experiments with the *luxCDABE* reporter were carried out by RG at the laboratory of Miguel Cámara (CBS, University of Nottingham, UK), with the valuable advice of Stephan Heeb and Manuel Romero.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00539/full#supplementary-material>

**DATASET S1 | Genes found in the transcriptomic analysis either induced or repressed in response to the expression of WH1(A31V)-mCh.** Genes were annotated (see **Figure 3**) as either induced ( $> +2.0$  fold ratio, in red) or repressed ( $< -2.0$ , in green) according to the ratios between the levels in bacteria expressing this hyper-amyloidogenic variant of the prionoid and those found in cells expressing WH1( $\Delta N37$ )-mCh. Each sheet displays a different point in the experimental time course and includes, for every entry, its probe ID, probability values, false discovery rates (FDR;  $\leq 0.05$ ) and functional annotations. Statistical significance was determined on three biological replicates.

**DATASET S2 | Proteins found co-aggregated with WH1(A31V)-mCh.** Each sheet displays the proteins identified as enriched in the aggregates formed by this hyper-amyloidogenic variant of the prionoid, but not by WH1( $\Delta N37$ )-mCh, at a different point in the experimental time course and in an independent biological replica (see **Figure 4**). For every entry, its reference number in the UniProt database and a functional description are displayed. Mass spectrometry parameters such as the calculated score, the % of coverage of the protein sequence by the identified peptides and their numbers, and the peptide spectrum match (PSM) value are also shown, together with the number of amino acids residues and the theoretical molecular weight of the targets.

## REFERENCES

- Aguzzi, A. (2009). Beyond the prion principle. *Nature* 459, 924–925. doi: 10.1038/459924a
- Al Mamun, A. A. M., Lombardo, M. J., Shee, C., Lisewski, A. M., Gonzalez, C., Lin, D., et al. (2012). Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science* 338, 1344–1348. doi: 10.1126/science.1226683
- Aslund, F., Zheng, M., Beckwith, J., and Storz, G. (1999). Regulation of the OxyR transcriptional factor by hydrogen peroxide and the cellular thiol-disulfide status. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6161–6165. doi: 10.1073/pnas.96.11.6161
- Barderas, R., Mendes, M., Torres, S., Bartolomé, R. A., López-Lucendo, M., Villar-Vázquez, R., et al. (2013). In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. *Mol. Cell. Proteomics* 12, 1602–1620. doi: 10.1074/mcp.M112.022848
- Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.* 65, 189–213. doi: 10.1146/annurev-micro-090110-102946
- Becher, A., and Schweizer, P. (2000). Integration-proficient *Pseudomonas aeruginosa* vectors for isolation of single-copy chromosomal *lacZ* and *lux* gene fusions. *Biotechniques* 29, 948–952. doi: 10.2144/111200024
- Bury-Moné, S., Noname, Y., Reymond, N., Barbet, R., Jacquet, E., Imbeaud, S., et al. (2009). Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. *PLoS Genet.* 5:e1000651. doi: 10.1371/journal.pgen.1000651
- Butterfield, S. M., and Lashuel, H. S. (2010). Amyloidogenic protein-membrane interactions: mechanistic insight from model systems. *Angew. Chem. Int. Ed. Engl.* 49, 5628–5654. doi: 10.1002/anie.200906670
- Capstick, D. S., Jomaa, A., Hanke, C., Ortega, J., and Elliot, M. A. (2011). Dual amyloid domains promote differential functioning of the chaplin proteins during *Streptomyces* aerial morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9821–9826. doi: 10.1073/pnas.1018715108
- Chapman, M. R., Robinson, L. S., Pinkner, J. S., Roth, R., Heuser, J., Hammar, M., et al. (2002). Role of *Escherichia coli* *curli* operons in directing amyloid fiber formation. *Science* 295, 851–855. doi: 10.1126/science.1067484
- Choi, H., Yang, Z., and Weisshaar, J. C. (2015). Single-cell, real-time detection of oxidative stress induced in *Escherichia coli* by the antimicrobial peptide CM15. *Proc. Natl. Acad. Sci. U.S.A.* 112, E303–E310. doi: 10.1073/pnas.1417703112
- Conchillo-Solé, O., de Groot, N. S., Avilés, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: A server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* 8:65. doi: 10.1186/1471-2105-8-65
- Costa, V., and Scorrano, L. (2012). Shaping the role of mitochondria in the pathogenesis of Huntington’s disease. *EMBO J.* 31, 1853–1864. doi: 10.1038/emboj.2012.65
- Eisenberg, D., and Jucker, M. (2012). The amyloid state of proteins in human diseases. *Cell* 148, 1188–1203. doi: 10.1016/j.cell.2012.02.022
- Eng, J. K., McCormack, A. L., and Yates, J. R. I. I. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein

- database. *J. Am. Soc. Mass. Spectrom.* 5, 976–989. doi: 10.1016/1044-0305(94)80016-2
- Felpeto-Santero, C., Rojas, A., Tortajada, M., Galán, B., Ramón, D., and García, J. L. (2015). Engineering alternative isobutanol production platforms. *AMB Express*. 5, 119. doi: 10.1186/s13568-015-0119-2
- Fernández, C., González-Rubio, G., Langer, J., Tardajos, G., Liz-Marzán, L. M., Giraldo, R., et al. (2016a). Nucleation of amyloid oligomers by RepA-WH1 prionoid-functionalized gold nanorods. *Angew. Chem. Int. Ed. Engl.* 55, 11237–11241. doi: 10.1002/anie.201604970
- Fernández, C., Núñez-Ramírez, R., Jiménez, M., Rivas, G., and Giraldo, R. (2016b). RepA-WH1, the agent of an amyloid proteinopathy in bacteria, builds oligomeric pores through lipid vesicles. *Sci. Rep.* 6:23144. doi: 10.1038/srep23144
- Fernández-Tresguerres, M. E., Moreno-Díaz de la Espina, S., Gasset-Rosa, F., and Giraldo, R. (2010). A DNA-promoted amyloid proteinopathy in *Escherichia coli*. *Mol. Microbiol.* 77, 1456–1469. doi: 10.1111/j.1365-2958.2010.07299.x
- Folsom, J. P., Parker, A. E., and Carlson, R. P. (2014). Physiological and proteomic analysis of *Escherichia coli* iron-limited chemostat growth. *J. Bacteriol.* 196, 2748–2761. doi: 10.1128/JB.01606-14
- Förster, A. H., and Gescher, J. (2014). Metabolic engineering of *Escherichia coli* for production of mixed-acid fermentation end-products. *Front. Bioeng. Biotechnol.* 2:16. doi: 10.3389/fbioe.2014.00016
- Foti, J. J., Devadoss, B., Winkler, J. A., Collins, J. J., and Walker, G. C. (2012). Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science* 336, 315–319. doi: 10.1126/science.1219192
- Frawley, E. R., and Fang, F. C. (2014). The ins and outs of bacterial iron metabolism. *Mol. Microbiol.* 93, 609–616. doi: 10.1111/mmi.12709
- García-Escudero, V., Martín-Maestro, P., Perry, G., and Avila, J. (2013). Deconstructing mitochondrial dysfunction in Alzheimer disease. *Oxid. Med. Cell. Longev.* 2013:162152. doi: 10.1155/2013/162152
- Gasset-Rosa, F., Coquel, A. S., Moreno-del Álamo, M., Chen, P., Song, X., Serrano, A. M., et al. (2014). Direct assessment in bacteria of prionoid propagation and phenotype selection by Hsp70 chaperone. *Mol. Microbiol.* 91, 1070–1087. doi: 10.1111/mmi.12518
- Gasset-Rosa, F., Maté, M. J., Dávila-Fajardo, C., Bravo, J., and Giraldo, R. (2008). Binding of sulphonated indigo derivatives to RepA-WH1 inhibits DNA-induced protein amyloidogenesis. *Nucleic Acids Res.* 36, 2249–2256. doi: 10.1093/nar/gkn067
- Giraldo, R. (2007). Defined DNA sequences promote the assembly of a bacterial protein into distinct amyloid nanostructures. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17388–17393. doi: 10.1073/pnas.0702006104
- Giraldo, R., Fernández, C., Moreno-del Álamo, M., Molina-García, L., Revilla-García, A., Sánchez-Martínez, M. C., et al. (2016). RepA-WH1 prionoid: clues from bacteria on factors governing phase transitions in amyloidogenesis. *Prion* 10, 41–49. doi: 10.1080/19336896.2015.1129479
- Giraldo, R., Fernández-Tornero, C., Evans, P. R., Díaz-Orejas, R., and Romero, A. (2003). A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain. *Nat. Struct. Biol.* 10, 565–571. doi: 10.1038/nsb937
- Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* 4:a011403. doi: 10.1101/cshperspect.a011403
- Haelterman, N. A., Yoon, W. H., Sandoval, H., Jaiswal, M., Shulman, J. M., and Bellen, H. J. (2014). A mitocentric view of Parkinson's disease. *Annu. Rev. Neurosci.* 37, 137–159. doi: 10.1146/annurev-neuro-071013-014317
- Harms, A., Maisonneuve, E., and Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* 354:aaf4268. doi: 10.1126/science.aaf4268
- Hipp, M. S., Park, S. H., and Hartl, F. U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* 24, 506–514. doi: 10.1016/j.tcb.2014.05.003
- Honn, M., Lindgren, H., and Sjøstedt, A. (2012). The role of MglA for adaptation to oxidative stress of *Francisella tularensis* LVS. *BMC Microbiol.* 12:14. doi: 10.1186/1471-2180-12-14
- Hosp, F., Vossfeldt, H., Heinig, M., Vasiljevic, D., Arumughan, A., Wyler, E., et al. (2015). Quantitative interaction proteomics of neurodegenerative disease proteins. *Cell Rep.* 11, 1134–1146. doi: 10.1016/j.celrep.2015.04.030
- Hwang, D., Lee, I. Y., Yoo, H., Gehlenborg, N., Cho, J. H., Petritis, B., et al. (2009). A systems approach to prion disease. *Mol. Syst. Biol.* 5, 252. doi: 10.1038/msb.2009.10
- Imlay, J. A. (2008). Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* 77, 755–776. doi: 10.1146/annurev.biochem.77.061606.161055
- Imlay, J. A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat. Rev. Microbiol.* 11, 443–454. doi: 10.1038/nrmicro3032
- Imlay, J. A. (2015a). Diagnosing oxidative stress in bacteria: not so easy as you might think. *Curr. Opin. Microbiol.* 24, 124–131. doi: 10.1016/j.mib.2015.01.004
- Imlay, J. A. (2015b). Transcription factors that defend bacteria against reactive oxygen species. *Annu. Rev. Microbiol.* 69, 93–108. doi: 10.1146/annurev-micro-091014-104322
- Jang, S., and Imlay, J. A. (2010). Hydrogen peroxide inactivates *Escherichia coli* Isc iron-sulphur assembly system, and OxyR induces the Suf system to compensate. *Mol. Microbiol.* 78, 1448–1467. doi: 10.1111/j.1365-2958.2010.07418.x
- Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., et al. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41, D605–D612. doi: 10.1093/nar/gks1027
- Lamark, T., Røkenes, T. P., McDougall, J., and Strøm, A. R. (1996). The complex *bet* promoters of *Escherichia coli*: regulation by oxygen (ArcA), choline (BetI), and osmotic stress. *J. Bacteriol.* 178, 1655–1662. doi: 10.1128/jb.178.6.1655-1662.1996
- Lee, J., Hiibel, S. R., Reardon, K. F., and Wood, T. K. (2009). Identification of stress-related proteins in *Escherichia coli* using the pollutant cis-dichloroethylene. *J. Appl. Microbiol.* 108, 2088–2102. doi: 10.1111/j.1365-2672.2009.04611.x
- Liebmans, S. W., and Chernoff, Y. O. (2012). Prions in yeast. *Genetics* 191, 1041–1072. doi: 10.1534/genetics.111.137760
- Lin, M. T., and Beal, M. F. (2006). Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* 443, 787–795. doi: 10.1038/nature05292
- Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J., and Taddei, F. (2008). Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3076–3081. doi: 10.1073/pnas.0708931105
- Liu, L., Zhang, K., Sandoval, H., Yamamoto, S., Jaiswal, M., Sanz, E., et al. (2015). Glial lipid droplets and ROS induced by mitochondrial defects promote neurodegeneration. *Cell* 160, 177–190. doi: 10.1016/j.cell.2014.12.019
- Malone, A. S., Chung, Y. K., and Yousef, A. E. (2006). Genes of *Escherichia coli* O157:H7 that are involved in high-pressure resistance. *Appl. Environ. Microbiol.* 72, 2661–2671. doi: 10.1128/AEM.72.4.2661-2671.2006
- Massé, E., Vanderpool, C. K., and Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J. Bacteriol.* 187, 6962–6971. doi: 10.1128/JB.187.20.6962-6971.2005
- Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., et al. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* 7, 237–242. doi: 10.1038/nmeth.1432
- Medina, I., Carbonell, J., Pulido, L., Madeira, S. C., Goetz, S., Conesa, A., et al. (2010). Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res.* 38, W210–W213. doi: 10.1093/nar/gkq388
- Messner, K. R., and Imlay, J. A. (1999). The identification of the primary sites of superoxide and hydrogen peroxide formation in the aerobic respiratory chain and sulphite reductase complex of *Escherichia coli*. *J. Biol. Chem.* 274, 10119–10128. doi: 10.1074/jbc.274.15.10119
- Molina-García, L., Gasset-Rosa, F., Moreno-del Álamo, M., Fernández-Tresguerres, M. E., Moreno-Díaz de la Espina, S., Lurz, R., et al. (2016). Functional amyloids as inhibitors of plasmid DNA replication. *Sci. Rep.* 6:25425. doi: 10.1038/srep25425
- Molina-García, L., and Giraldo, R. (2014). Aggregation interplay between variants of the RepA-WH1 prionoid in *Escherichia coli*. *J. Bacteriol.* 196, 2536–2542. doi: 10.1128/JB.01527-14
- Moreno-del Álamo, M., Moreno-Díaz de la Espina, S., Fernández-Tresguerres, M. E., and Giraldo, R. (2015). Pre-amyloid oligomers of the proteotoxic RepA-WH1 prionoid assemble at the bacterial nucleoid. *Sci. Rep.* 5:14669. doi: 10.1038/srep14669



- Myers, K. S., Yan, H., Ong, I. M., Chung, D., Liang, K., Tran, F., et al. (2013). Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. *PLoS Genet.* 9:e1003565. doi: 10.1371/journal.pgen.1003565
- Narayan, P., Ehsani, S., and Lindquist, S. (2014). Combating neurodegenerative disease with chemical probes and model systems. *Nat. Chem. Biol.* 10, 911–920. doi: 10.1038/nchembio.1663
- Olzsch, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G., et al. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* 144, 67–78. doi: 10.1016/j.cell.2010.11.050
- Pallarés, I., Iglesias, V., and Ventura, S. (2016). The Rho termination factor of *Clostridium botulinum* contains a prion-like domain with a highly amyloidogenic core. *Front. Microbiol.* 6:1516. doi: 10.3389/fmicb.2015.01516
- Pósfai, G., Plunkett, G., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046. doi: 10.1126/science.1126439
- Riek, R., and Eisenberg, D. A. (2016). The activities of amyloids from a structural perspective. *Nature* 539, 227–235. doi: 10.1038/nature20416
- Romero, D., Aguilar, C., Losick, R., and Kolter, R. (2010). Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2230–2234. doi: 10.1073/pnas.0910560107
- Sánchez-Gorostiaga, A., Palacios, P., Martínez-Arteaga, R., Sánchez, M., Casanova, M., and Vicente, M. (2016). Life without division: physiology of *Escherichia coli* FtsZ-deprived filaments. *mBio* 7:e01620-16. doi: 10.1128/mBio.01620-16
- Seaver, L. C., and Imlay, J. A. (2004). Are respiratory enzymes the primary sure of intracellular hydrogen peroxide? *J. Biol. Chem.* 279, 48742–48750. doi: 10.1074/jbc.M408754200
- Seo, S. W., Kim, D., Szubin, R., and Palsson, B. O. (2015). Genome-wide reconstruction of OxyR and SoxRS transcriptional regulatory networks under oxidative stress in *Escherichia coli* K-12 MG1655. *Cell Rep.* 12, 1289–1299. doi: 10.1016/j.celrep.2015.07.043
- Sivanathan, V., and Hochschild, A. (2012). Generating extracellular amyloid aggregates using *E. coli* cells. *Genes Dev.* 26, 2659–2667. doi: 10.1101/gad.205310.112
- Sorce, S., Nuvolone, M., Keller, A., Falsig, J., Varol, A., Schwarz, P., et al. (2014). The role of the NADPH oxidase NOX2 in prion pathogenesis. *PLoS Pathog.* 10:e1004531. doi: 10.1371/journal.ppat.1004531
- Spivak, M., Weston, J., Bottou, L., Käll, L., and Noble, W. S. (2009). Improvements to the Percolator algorithm for peptide identification from shotgun proteomics data sets. *J. Proteome Res.* 8, 3737–3745. doi: 10.1021/pr801109k
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Taylor, J. P., Brown, R. H. Jr., and Cleveland, D. W. (2016). Decoding ALS: from genes to mechanism. *Nature* 539, 197–206. doi: 10.1038/nature20413
- Torreira, E., Moreno-del Álamo, M., Fuentes-Perez, M. E., Fernández, C., Martín-Benito, J., Moreno-Herrero, F., et al. (2015). Amyloidogenesis of the bacterial prionoid RepA-WH1 recapitulates dimer to monomer transitions of RepA in DNA replication initiation. *Structure* 23, 183–189. doi: 10.1016/j.str.2014.11.007
- Uden, G., and Bongaerts, J. (1997). Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim. Biophys. Acta* 1320, 217–234. doi: 10.1016/S0005-2728(97)00034-0
- Van Gerven, N., Klein, R. D., Hultgren, S. J., and Remaut, H. (2015). Bacterial amyloid formation: structural insights into curli biogenesis. *Trends Microbiol.* 23, 693–706. doi: 10.1016/j.tim.2015.07.010
- Weber, H., Polen, T., Heuveling, J., Wendisch, V. F., and Hengge, R. (2005). Genome-wide analysis of the general stress response network in *Escherichia coli*:  $\sigma$ S-dependent genes, promoters and sigma factor selectivity. *J. Bacteriol.* 187, 1591–1603. doi: 10.1128/JB.187.5.1591-1603.2005
- Winkler, J., Seybert, A., König, L., Pruggnaller, S., Haselmann, U., Sourjik, V., et al. (2010). Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J.* 29, 910–923. doi: 10.1038/emboj.2009.412
- Woerner, A. C., Frottin, F., Hornburg, D., Feng, L. R., Meissner, F., Patra, M., et al. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science* 351, 173–176. doi: 10.1126/science.aad2033
- Woodmansee, A. N., and Imlay, J. A. (2002). Reduced flavins promote oxidative DNA damage in non-respiring *Escherichia coli* by delivering electrons to intracellular free iron. *J. Biol. Chem.* 277, 34055–34066. doi: 10.1074/jbc.M203977200
- Yuan, A. H., and Hochschild, A. (2017). A bacterial global regulator forms a prion. *Science* 355, 198–201. doi: 10.1126/science.aai7776

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Citation:** Molina-García L, Moreno-del Álamo M, Botías P, Martín-Moldes Z, Fernández M, Sánchez-Gorostiaga A, Alonso-del Valle A, Nogales J, García-Cantalejo J and Giraldo R (2017) Outlining Core Pathways of Amyloid Toxicity in Bacteria with the RepA-WH1 Prionoid. *Front. Microbiol.* 8:539. doi: 10.3389/fmicb.2017.00539

Copyright © 2017 Molina-García, Moreno-del Álamo, Botías, Martín-Moldes, Fernández, Sánchez-Gorostiaga, Alonso-del Valle, Nogales, García-Cantalejo and Giraldo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Cadaver Thanatobiome Signatures: The Ubiquitous Nature of *Clostridium* Species in Human Decomposition

Gulnaz T. Javan<sup>1\*</sup>, Sheree J. Finley<sup>2</sup>, Tasia Smith<sup>1</sup>, Joselyn Miller<sup>1</sup> and Jeremy E. Wilkinson<sup>3</sup>

<sup>1</sup> Forensic Science Program, Physical Sciences Department, Alabama State University, Montgomery, AL, United States,

<sup>2</sup> Physical Sciences Department, Alabama State University, Montgomery, AL, United States, <sup>3</sup> Research and Testing Laboratory, RTL Genomics, Lubbock, TX, United States

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
United States

### Reviewed by:

Antonio González-Martín,  
Complutense University of Madrid,  
Spain  
Miguel Angel Cevallos,  
Universidad Nacional Autónoma  
de México, Mexico

### \*Correspondence:

Gulnaz T. Javan  
gjavan@alasu.edu

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 22 August 2017

**Accepted:** 12 October 2017

**Published:** 30 October 2017

### Citation:

Javan GT, Finley SJ, Smith T, Miller J  
and Wilkinson JE (2017) Cadaver  
Thanatobiome Signatures:  
The Ubiquitous Nature of *Clostridium*  
Species in Human Decomposition.  
Front. Microbiol. 8:2096.  
doi: 10.3389/fmicb.2017.02096

Human thanatobiome studies have established that an abundant number of putrefactive bacteria within internal organs of decaying bodies are obligate anaerobes, *Clostridium* spp. These microorganisms have been implicated as etiological agents in potentially life-threatening infections; notwithstanding, the scale and trajectory of these microbes after death have not been elucidated. We performed phylogenetic surveys of thanatobiome signatures of cadavers' internal organs to compare the microbial diversity between the 16S rRNA gene V4 hypervariable region and V3-4 conjoined regions from livers and spleens of 45 cadavers undergoing forensic microbiological studies. Phylogenetic analyses of 16S rRNA gene sequences revealed that the V4 region had a significantly higher mean Chao1 richness within the total microbiome data. Permutational multivariate analysis of variance statistical tests, based on unweighted UniFrac distances, demonstrated that taxa compositions were significantly different between V4 and V3-4 hypervariable regions ( $p < 0.001$ ). Of note, we present the first study, using the largest cohort of criminal cases to date, that two hypervariable regions show discriminatory power for human postmortem microbial diversity. In conclusion, here we propose the impact of hypervariable region selection for the 16S rRNA gene in differentiating thanatobiomic profiles to provide empirical data to explain a unique concept, the Postmortem *Clostridium* Effect.

**Keywords:** thanatobiome, *Clostridium*, 16S rRNA gene, V4 hypervariable regions, V3-4 hypervariable regions, Postmortem *Clostridium* Effect

## INTRODUCTION

Thanatobiome studies have determined that there is extremely rapid postmortem overgrowth of *Clostridium* spp. within decaying internal body sites (e.g., blood, bone marrow, liver, prostate) (Clement et al., 2016; Javan et al., 2016a,b, Adserias-Garriga et al., 2017a,b; Thomas et al., 2017; Zhao et al., 2017). Human decomposition is a multifactorial process mediated by

microbes, which inhabit, proliferate, and die externally and internally throughout dead biomass (Javan et al., 2016a,b). *Clostridium* spp. are strict anaerobes and common symbiotic bacteria located in healthy intestines. High abundance of nine *Clostridium* spp., namely *C. sordellii*, *C. difficile*, *C. bartlettii*, *C. bifermentans*, *C. limosum*, *C. haemolyticum*, *C. botulinum*, and *C. novyi*, were discovered by next-generation sequencing of 16S rRNA gene amplicons in previous thanatomicrobiome studies of human postmortem samples (Can et al., 2014; Javan et al., 2016b). Enteric bacteria, including *Clostridium* spp., are capable of translocating to surrounding tissues within 5–48 h after death at 25°C (Morris et al., 2006). *Clostridium* spp. reside in the mucosal layer and the intestinal epithelial monolayer, and they metabolize predigested hexoses entering from the stomach to acetic acid, acetone, butanoic acid, butanol, and ethanol which bacteria then ferment to pyruvate (Corry, 1978; Boumba et al., 2008; Janaway et al., 2009). Furthermore, Clostridia break down the amino acid threonine to propanol using threonine dehydratase,  $\alpha$ -ketobutyrate synthase, and NAD-linked propanol dehydrogenase (Boumba et al., 2008).

Prokaryotic 16S rRNA gene amplicon sequences are extensively used in forensic microbiology as reliable biomarkers for taxonomic classification and phylogenetic analysis of the microbiome of death. The thanatomicrobiome, which is defined as microbial succession in decomposing remains (e.g., blood, bone marrow, liver, reproductive organs), can provide evidence concerning interactions between microorganisms and their mammalian hosts. Microbes symbiotically cohabitate with humans during life, but they also participate in the nature and trajectory of decomposition. The host's death introduces chaos in microbial communities as the body becomes an abounding source of nutrients (Mondor et al., 2012). The question then arises, “What hypervariable region(s) of the 16S rRNA gene best profile the shifts that occur in response to the massive proliferation of microbiota after death?” The phenomenon that these signatures are left behind by the corpse provides unique forensic potential to make available trace evidence that can be used in microbial forensics.

Analysis of the very informative 16S rRNA gene is commonly used as a genetic marker for profiling prokaryotic communities (Lane et al., 1985; Woese et al., 1990; Baker et al., 2003; Tringe and Hugenholtz, 2008; Wang and Qian, 2009). Recent postmortem microbiome studies have focused on the 16S rRNA gene Class I, which spans the V4 region (Hyde et al., 2013, 2015; Damann et al., 2015; Javan et al., 2016b; Metcalf et al., 2016). The V4 hypervariable region is one of the major functional parts of the microbial gene because it encompasses a portion of the “690 hairpin” (Morosyuk et al., 2000; Wimberly et al., 2000) and decoding center (Schluenzen et al., 2000; Morosyuk et al., 2001). The V3 region is categorized in Class II, which is peripheral to the two functional centers of the 16S rRNA gene (Schluenzen et al., 2000; Schuwirth et al., 2005). Studies have shown that V4 is the best region for phylogenetic studies, particularly at the phylum level (Yang et al., 2016).

A key question is, which sub-region (V4 or V3-4) is more effective for phylogenetic studies of the human thanatomicrobiome? To explore the potential to determine

cadaver thanatomicrobiome signatures using two hypervariable regions of the 16S rRNA gene, we compared the performance of primers 515F-806R (V4) to 357wF-785R (conjoined V3-4) hypervariable regions (Johnson et al., 2016). We hypothesized that by modulating 16S rRNA gene hypervariable fragment lengths on the Illumina MiSeq platform, the two specified regions would produce dissimilar microbial signatures.

Bioinformatic surveys have shown that hypervariable regions of 16S rRNA gene differ in the detection of sequence diversity; thus, a particular region may function well for ascertaining a spectrum of bacterial taxa whereas a different region may exhibit a distinct degree of taxonomic diversity. The V3-4 amplicon has a read length of twice 250 bp that offers an ideal target for Illumina paired-end sequencing and will provide a suitable framework for V4 region comparisons of the effectiveness of hypervariable region performance. Here, we performed a phylogenetic assessment of species distinctions using V4 versus V3-4 hypervariable regions from postmortem liver and spleen samples from criminal cases. Furthermore, we determined for the first time that fast-growing members of postmortem microbial communities, *Clostridium* spp., that usually predominate at longer PMIs, also are the most prominent prokaryotes even at shorter time intervals (PMI = 4 h).

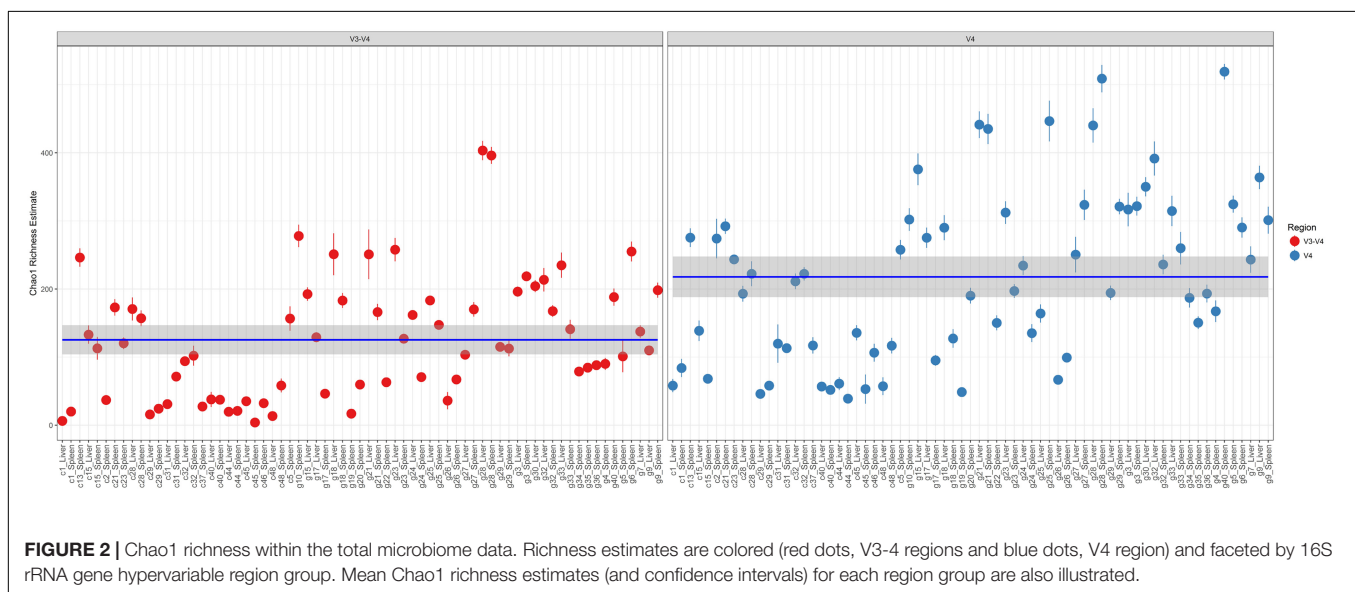
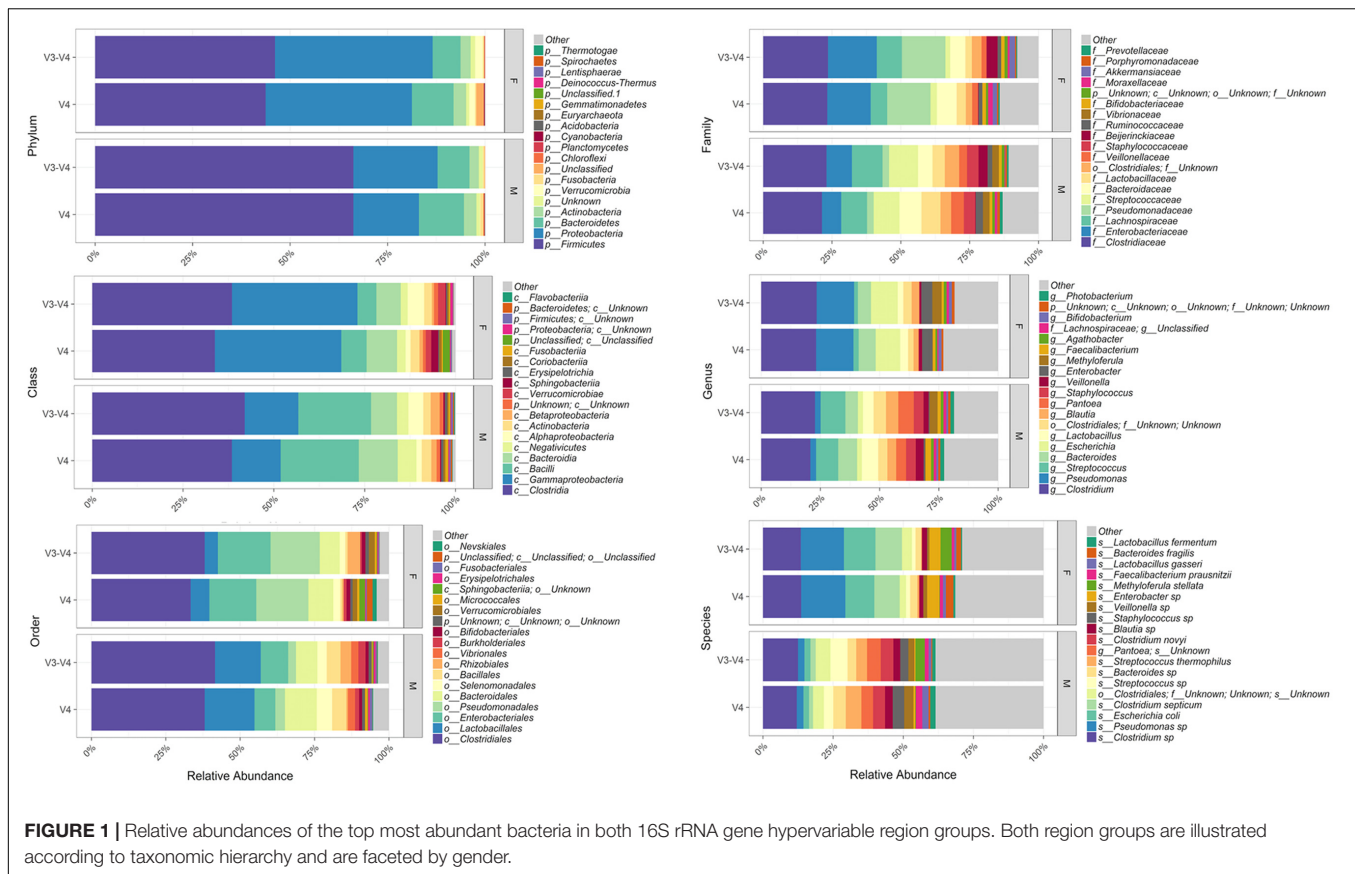
## MATERIALS AND METHODS

### Postmortem Sampling of Human Corpses

Postmortem samples included 28 male and 17 female corpses from the Alabama Department of Forensic Sciences in Montgomery, AL and The Office of the District One Medical Examiner in Pensacola, FL, United States. Demographic data were collected on each of the 45 corpses (i.e., age, gender, ethnicity, cause of death, PMI) (**Supplementary Table S1**). The study was approved by Alabama State University Institutional Review Board (IRB) number 2016011. Time of death of each corpse was certified from official Daily Crime Logs created by local police departments. Bodies were stored in the morgues at 1°C until time of tissue dissection. Approximately 10 mg of liver and spleen tissues were dissected using sterile scalpels and placed in polyethylene bags in an examination area at 20°C. Organs were transported on dry ice to the Thanatos Laboratory at Alabama State University. Specimens were stored at –80°C until time of DNA extraction.

### DNA Extraction of Postmortem Samples

Approximately 10 mg of thawed liver and spleen tissues were placed into Lysing matrix E tubes (MP Biomedicals) containing zirconia and silica beads, 0.5 ml phenol/chloroform/isoamyl alcohol (25:24:1) (TE saturated, pH 8.0) and 0.5 ml of 2× TENS buffer [100 mM Tris-HCl (pH 8.0), 40 mM EDTA, 200 mM NaCl, 2% SDS] (Wan et al., 2011). Tubes were homogenized by mechanical horizontal vortexing in a Mini Beadbeater (BioSpec Products) at speed 40 and time 6, briefly



cooled on ice, and centrifuged at 16,000 rpm for 5 min. Supernatants were transferred to 2.0 ml Phase Lock Gel tubes (Invitrogen) containing 0.3 ml of 7.5 M ammonium acetate and equal volumes of chloroform. Tubes were mixed by repeated moderate inverting 10 times and supernatants were transferred into new tubes containing 0.6 volumes of ice

cold isopropanol and 3  $\mu$ l of GlycoBlue Coprecipitant (Life Technologies). After gently inverting several times, samples were incubated at  $-80^{\circ}\text{C}$  for 10 min. Following centrifugation at 16,000 rpm for 5 min, isopropanol was decanted and pellets were washed with cold 80% ethanol and allowed to dry for 5 min. Pellets were eluted with 100  $\mu$ l of TE buffer. DNA was



quantified by NanoDrop2000™ (Thermo Scientific) measuring the absorbance at 260 nm.

### Illumina MiSeq Sequencing

V4 and V3-4 hypervariable regions of 16S rRNA gene were amplified for sequencing at RTL Genomics (Research and Testing Laboratory, Lubbock, TX, United States) in two-step, independent reactions using HotStar Taq Master Mix Kit (Qiagen) with universal primers 515F-806R for the V4 region and primer constructs 357wF/785R for the longer, combined V3-4 regions. Primers for the first step were constructed using the fragment-specific forward and reverse primers (515F-806R or 357wF-785R) with the Illumina i5 and i7 sequencing primers added to the 5'-end of each, respectively. Products from the first amplification were added to a second PCR step based on qualitatively determined concentrations (amplicons were run on a 2% ethidium gel, gel bands were scored, and a volume of products was added to the second PCR based on the scores). Primers for the second PCR step were designed

TABLE 1 | Results of ANOVA, that tested for differences in Chao1 richness.

	df	Sum Sq.	Mean Sq.	F-value	Pr (>F)
Region	1	294,432.16	294,432.16	53.02	0
Organ	1	9994.93	9994.93	1.8	1.822
Gender	1	109,506.49	109,506.49	19.72	0
Manner of death	4	555,725.31	138,931.33	25.02	0
PMI	1	21.15	21.15	0	0.9509
Season	3	73,761.48	24,587.16	4.43	0.0054
Weight	1	16.79	16.79	0	0.9562
Height	1	48,683.31	48,683.31	8.77	0.0037
Residuals	123	682,989.39	5552.76		

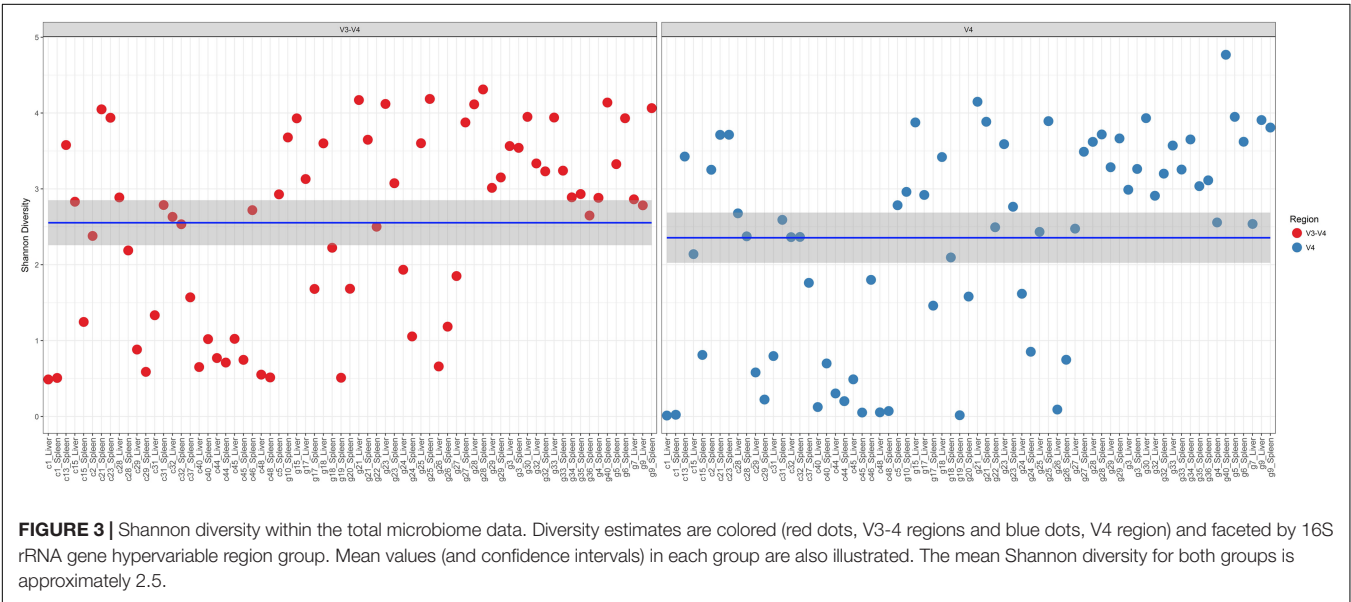
Degrees of freedom (df) corresponds to one less than the number of values in the set of means. The p-values are derived from the F-distribution and the significance level is Pr (>F) < 0.001 (shaded cells).

TABLE 2 | Results of ANOVA, that tested for differences in Shannon diversity.

	df	Sum Sq.	Mean Sq.	F-value	Pr (>F)
Region	1	1.34	1.34	1.69	0.196
Organ	1	0.29	0.29	0.37	0.5467
Gender	1	13.67	13.67	17.21	0.0001
Manner of death	4	62.8	15.7	19.76	0
PMI	1	3.98	3.98	5.01	0.027
Season	3	15.73	5.24	6.6	0.0004
Weight	1	0.2	0.2	0.25	0.6153
Height	1	8.56	8.56	10.78	0.0013
Residuals	123	97.74	0.79		

Degrees of freedom (df) corresponds to one less than the number of values in the set of means. The p-values are derived from the F-distribution and the significance level is Pr (>F) < 0.001 (shaded cells).

using Illumina Nextera PCR primers with 8 bp dual indexes. Each PCR amplification included 9 µl of sterile deionized H<sub>2</sub>O, 0.5 µl of 5 µM forward primer, 0.5 µl of 5 mM reverse primer, 1 µl of DNA template, and 14 µl of Taq Master Mix. The negative control was a reaction mixture with no template DNA. PCR reaction conditions included initial denaturation at 95°C for 5 min, then 25 cycles of 94°C for 30 s, annealing at 54°C for 40 s, and extension at 72°C for 1 min, followed by 1 cycle of 72°C for 10 min and 4°C hold. Barcoding PCR reactions were conducted under the same conditions, except with only 10 cycle extensions. Amplification products were visualized with eGels (Life Technologies). Products were then pooled equimolar and each pool was size selected in two rounds using SPRIselect beads (BeckmanCoulter) in a 0.7 ratio for both rounds. Size selected pools were then quantified using Qubit 2.0 fluorometer (Life Technologies) and loaded on an Illumina MiSeq 2x300 flow cell at 10 pM and sequenced.



## Bioinformatic Analysis

The sequence data were analyzed using a standard microbial diversity analysis pipeline, which consisted of two major stages, denoising and chimera detection followed by microbial diversity analysis. Denoising was performed using various techniques to remove short sequences, singleton sequences, and noisy reads. Chimera detection was performed using the UCHIME chimera detection software in *de novo* mode (Edgar et al., 2011). Lastly, remaining sequences were then corrected base-by-base to help remove noise from within each sequence. During the diversity analysis stage, each sample was run through the analysis pipeline to cluster reads into OTUs (at 97% identity) using the UPARSE algorithm (Edgar, 2013), and then globally aligned using the USEARCH global algorithm (Edgar, 2010) against a database of high-quality 16S rRNA gene sequences to determine taxonomic classifications. After OTU selection was performed, a phylogenetic tree was constructed in Newick format from a multiple sequence alignment of OTUs done in MUSCLE (Edgar, 2004a,b) and generated in FastTree (Price et al., 2009, 2010; Shah et al., 2016).

Microbial diversity of cadaver samples was examined from two perspectives using the phyloseq package in R (McMurdie and Holmes, 2013). First, overall richness (i.e., number of distinct nucleic acid sequences present within the microbiome) was expressed as the number of OTUs and was quantified using the Chao1 richness estimator. Secondly, overall microbial diversity, determined by both richness and evenness and the distribution of abundance among distinct taxa, was expressed as Shannon–Wiener species diversity. Measures of microbial diversity were screened for group (region, organ, gender, manner of death, PMI, season, location, weight, and height) differences using an analysis of variance (ANOVA). Multivariate differences among groups were evaluated with permutational multivariate analysis of variance (PERMANOVA) using distance matrices function ADONIS (Oksanen, 2011). For PERMANOVA, ADONIS distances among samples first were calculated using unweighted or weighted UniFrac via the phyloseq package in R (McMurdie and Holmes, 2013), and then an ANOVA-like simulation was conducted

to test for group differences. Principal coordinates analysis (PCoA) using unweighted and weighted UniFrac distances and relative abundance bar plots were generated to visualize relationships and differences among groups. All analyses were conducted in R (R Development Core Team, 2010) and all plots were generated using the ggplot2 package (Wickham, 2009).

## RESULTS

### Thanatobiome Sequencing of Postmortem Liver and Spleen Samples

Bioinformatic characterization of relative abundances and microbial diversity of the thanatobiome was performed through metagenomic analyses in order to determine if there was greater discriminatory power exhibited by 16S rRNA gene V4 versus V3-4 hypervariable region amplicons. Operational taxonomic unit (OTU) data were validated by rarefaction analyses. Rarefaction data confirmed complete coverage until 20,000 sequences to observe all taxa as shown by convergence of vertical asymptotes for all curves (data not shown). Relative abundances of the top most abundant bacteria according to taxonomic hierarchy are shown in **Figure 1**. The highest percentage of bacteria on the order level was Clostridiales and seven of the top species were *Clostridium* spp. Furthermore, 95% of samples contained *Clostridium* spp., whereas six of the seven samples that did not contain *Clostridium* spp. were from the V3-4 hypervariable region.

### Thanatobiome Alpha and Beta Diversity Analyses

Comparison of Chao1 richness estimations was calculated and the V4 hypervariable region had a higher proportion of average calculated estimates than V3-4 (**Figure 2**). For V4 region amplicons, the average Chao1 richness estimate was 217 species, whereas V3-4 averaged 125 species. Also, ANOVA analysis revealed a statistically significant difference in Chao1 richness

**TABLE 3 |** Results of the permutational multivariate analysis of variance using distance matrices function ADONIS, unweighted and weighted UniFrac, respectively.

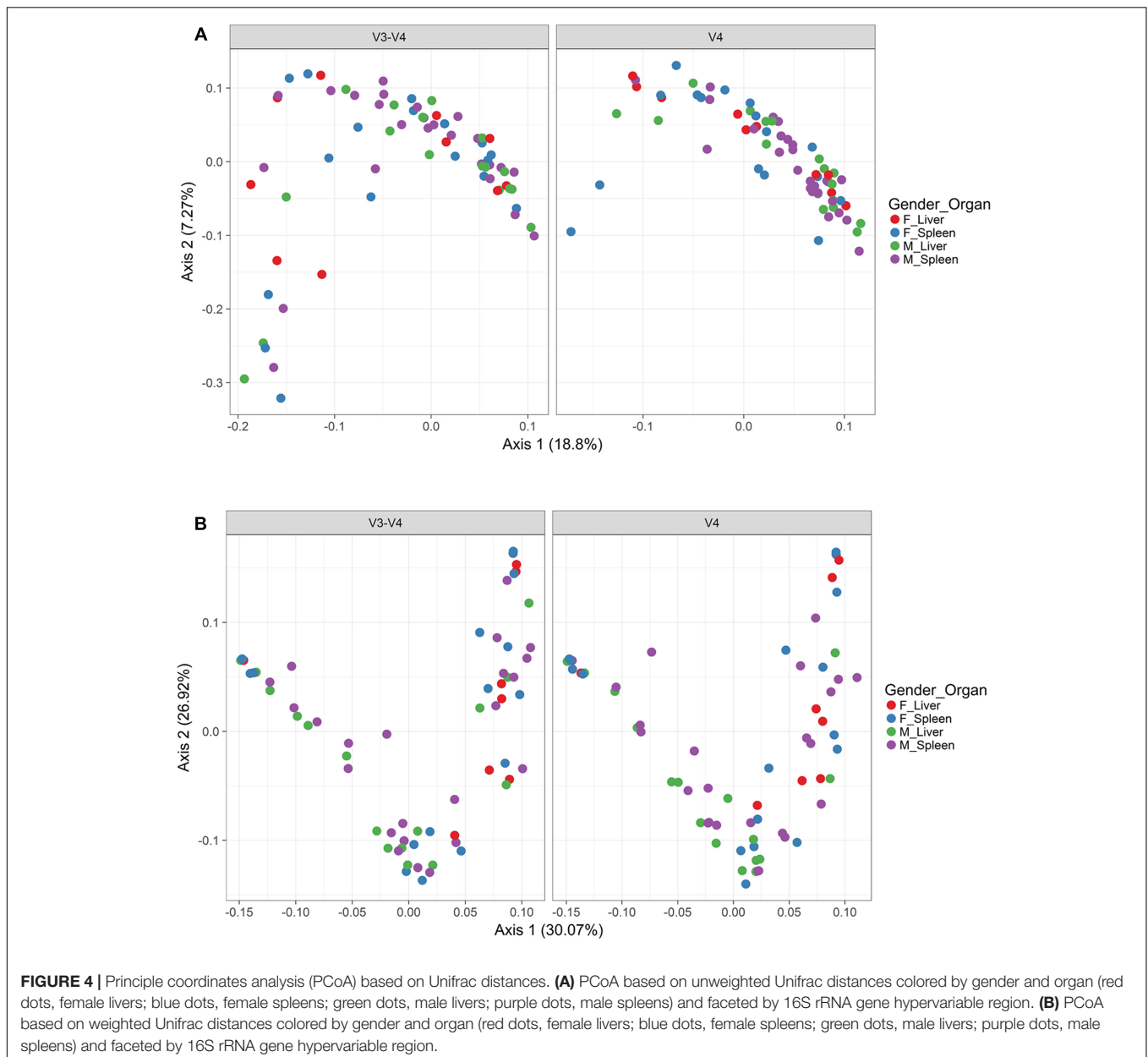
	Unweighted UniFrac						Weighted UniFrac					
	df	Sum Sq.	Mean Sq.	F-model	R <sup>2</sup>	Pr (>F)	df	Sum Sq.	Mean Sq.	F-model	R <sup>2</sup>	Pr (>F)
Region	1	1.5	1.5	8.35	0.05	0.001	1	0.15	0.15	1.26	0.01	0.26
Organ	1	0.19	0.19	1.06	0.01	0.284	1	0.08	0.08	0.65	0	0.644
Gender	1	0.49	0.49	2.72	0.02	0.002	1	0.91	0.91	7.71	0.04	0.001
Manner of death	4	2.52	2.52	3.51	0.08	0.001	4	2.04	0.51	4.29	0.09	0.001
PMI	1	0.33	0.33	1.85	0.01	0.025	1	0.62	0.62	5.19	0.03	0.002
Season	3	1.3	0.43	2.41	0.04	0.001	3	1.61	0.54	4.52	0.07	0.001
Residuals	123	22.07	0.18		0.71		123	14.6	0.12		0.67	
Total	137	29.78			1		137	21.87			1	

Degrees of freedom (df) corresponds to one less than the number of values in the set of means. The p-values are derived from the F-distribution and the significance level is  $Pr(>F) < 0.001$  (shaded cells).

between two 16S rRNA gene regions (ANOVA;  $p < 0.001$ ) (**Table 1**). In comparisons of gender and manner of death (accident, homicide, natural, suicide, undetermined), statistically significant differences were observed; however, patterns of species richness were statistically independent of time of death. Conversely, no significance was observed in Chao1 richness in the comparison of other variables (e.g., season of death).

In a comparison of the Shannon Diversity within the total microbiome data, both hypervariable regions demonstrated an overall similar profile (**Figure 3**). The average Shannon Diversity index representing both regions was approximately 2.63 for the V4 region and 2.55 for V3-4; however, significant differences were observed for gender, manner of death, and season of death (ANOVA;  $p < 0.001$ , **Table 2**).

Results of multivariate difference interactions between both 16S rRNA gene regions and other variables resulted in statistical significance in region and location (interaction ADONIS;  $p = 0.001$ ), which demonstrated that location was the only factor that had a confounding effect on region and that other factors were not confounding results. Results of ADONIS based on weighted UniFrac distances demonstrated significant differences among gender, manner of death, and season ( $p < 0.001$ ), but not between V4 and V3-4 16S rRNA gene regions (**Table 3**). Given the fact that significant differences were observed among regions in unweighted but not in weighted UniFrac, the presence or absence of OTUs was more dissimilar than the abundance of OTUs (**Table 3**).





**FIGURE 5 |** Principle Coordinates Analysis (PCoA) based on Unifrac distances. **(A)** PCoA based on unweighted Unifrac distances colored by manner and season (red dots, accident; blue dots, homicide; green dots, natural; purple dots, suicide; orange dots, undetermined). **(B)** PCoA based on weighted Unifrac distances colored by manner and season (red dots, accident; blue dots, homicide; green dots, natural; purple dots, suicide; orange dots, undetermined).

In order to visualize beta diversity differences between 16S rRNA gene regions, PCoA plots were generated based on unweighted (**Figure 4A**) and weighted Unifrac distances metrics (**Figure 4B**). For unweighted UniFrac, there was relatively low variance among two 16S rRNA gene regions, with only 18.10% of variance explained by primary Axis 1 and 7.27% explained by secondary Axis 2. PCoA plots based on unweighted and weighted Unifrac distances were generated faceted by manner of death and season (**Figures 5A,B**, respectively). Furthermore, samples for the V4 region clustered compactly together more than those for V3-4. For weighted UniFrac, there was more variance among 16S rRNA gene regions compared to unweighted UniFrac PCoA, with 30.07% of the variance explained by primary Axis 1 and 26.92% explained by secondary Axis 2.

## DISCUSSION

Human antemortem microbiotas are well documented by the Human Microbiome Project (Peterson et al., 2009), for various body locations of healthy individuals; however, currently there is

a paucity of knowledge and need for an in-depth interpretation concerning postmortem microbial communities of internal body sites. Our thanatomicrobiome research represents the largest exploratory study examining a cohort of 45 corpses in fresh and bloat stages. Also, the study provides the first extensive catalog of postmortem microbiomes obtained in internal locations analyzed by two different hypervariable regions of the 16S rRNA gene. Approximately 95% of the postmortem liver and spleen profiled in this study involved *Clostridium* spp. Moreover, the findings revealed that V4 and V3-4 hypervariable testings represent incongruent phylotype diversity and consequently support individual representative assessments of the thanatomicrobiome. For example, study-specific disparities were observed; *Clostridium* spp. were not obtained in only one of the V4 region sequences. On the other hand, these species were not obtained in six V3-4 region analyses. Here, we demonstrate that amplicons more sufficient to discriminate *Clostridium* spp. in postmortem tissue are derived from the V4 hypervariable region. According to previous thanatomicrobiome studies, *Clostridium* spp. predominated at long PMIs (up to 10 days) (Javan et al., 2016b). However, the



current study determined that these Gram-positive, anaerobic extremophiles also predominate at shorter PMIs (4 h).

Our results support Yang et al. (2016) study based on geodesic distances that suggested V4 was the best sub-region for phylogenetic analysis. In the present study, we confirmed that the V4 region, belonging to Class I, had elevated sensitivity for the detection of forensically relevant bacteria; whereas V3 from Class II showed moderate sensitivity. Of particular interest, there was a higher enrichment of the species *Methyloferula stellata* discovered by targeting V3-4 amplicons compared to V4 (Figure 1). *M. stellata* is a methanotroph that grows exclusively on methane and methanol (Gill and Landi, 2011; Vorobev et al., 2011). During the bloat stage of human decomposition, methane, along with various odoriferous putrefaction gases, is produced in high abundance by anaerobic fermentation especially emanating from the gastrointestinal tract (Gill and Landi, 2011). Our study is the first to confirm a bacterial taxon that thrives on one of the putrefying gasses produced during decomposition through the use of 16S rRNA gene V3-4 combined regions in human internal body sites. Another very interesting finding was high abundances of three bacteria, *Escherichia coli*, *C. septicum*, and a *Pseudomonas* sp. detected only in female cases using both hypervariable regions (Figure 1).

In the last decade, postmortem microbiology studies have created novel thanatomicrobiome and epinecrotic communities catalogs using expertise in genetics, next-generation sequencing, and bioinformatics. The creation of a Human Postmortem Microbiome Project (HPMP) will facilitate the development of *modus operandi* used to empower data comparisons obtained from different national and international laboratories. Extension of existing standard operating procedures that cover sampling, processing, sequencing, and analysis will conceive universal standards for microbial analysis to unify the global research community. In addition, the HPMP framework includes research emphases that will provide the scientific community with a hub through which researchers can explore microbial life after death.

## Postmortem Clostridium Effect

The current research defines a new scientific concept, the “Postmortem Clostridium Effect” (PCE), which refers to facultative anaerobic *Clostridium* spp. that are ubiquitous during human decomposition. There are three dynamics that contribute to *Clostridium* species’ omnipresence in decaying humans; one factor involves its very fast doubling time. For example, a species found in the present study, *C. perfringens*, has the most rapid generation time of approximately 7.4 min at optimal temperatures (37–45°C) (Willardsen et al., 1979). The second factor is the bacterium’s proteolytic functions. *Clostridium* spp. have collagenases that digest native vertebrate collagen fibers which confer the ability to breach colon epithelial surfaces and mucosal layers and transmigrate to proximate tissues (Harrington, 1996; Burcham et al., 2016). The last advantageous putrefactive factor of *Clostridium* spp. develops via the cessation of the human heart which results in hypoxia (Gevers, 1975; Proskuryakov et al., 2003). A corpse that lacks oxygenated blood can facilitate enteric anaerobic bacteria, naturally found in the colon (e.g., *Clostridium* spp.), to efficiently and rapidly flourish

in the nutrient-rich host. Previous human decomposition studies reported a marked shift from communities dominated by aerobic bacteria to anaerobic at the end of the bloat stage (Hyde et al., 2013). Taken together, these host–microbe factors within decaying biomass lead to the efficient functioning of the PCE.

## CONCLUSION

The thanatomicrobiome contributes a substantial function in modulating human decomposition. Studies are needed to elucidate if hypervariable regions are capable to discriminate all bacterial species; therefore, our emphasis was on the phylogenic resolution produced by small subunit surveys to characterize microbial mediators of decay. Conceivably, thanatomicrobiome analysis will be used to build predictive Thanatos models employing the PCE that can further designate the recovery of distinct community types associated with postmortem microbial communities.

## ETHICS STATEMENT

The study was approved by the Committee for the Protection of Human Subjects, Alabama State University Institutional Review Board (IRB) number 2016011. Methods were in accordance with relevant guidelines and regulations regarding working with cadavers. Written informed consent was obtained from next-of-kin relatives of the cases.

## AVAILABILITY OF DATA AND MATERIAL

All data generated or analyzed during this study are included in this published article and its supplementary information files.

## AUTHOR CONTRIBUTIONS

GJ designed the study and collected human corpses. GJ, SF, JM, and TS extracted genomic DNA, PCR, gel electrophoresis the samples. JW performed MiSeq sequencing and data analysis. GJ and SF wrote and edited the article. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the National Science Foundation (NSF) grant HRD 1401075.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02096/full#supplementary-material>

**TABLE S1 |** Metadata of 16S rRNA regions, age, gender, ethnicity, season, location, cause of death, PMI.

## REFERENCES

- Adserias-Garriga, J., Hernandez, M., Quijada, N. M., Lázaro, D. R., Steadman, D., and Garcia-Gil, J. (2017a). Daily thanatomicrobiome changes in soil as an approach of postmortem interval estimation: an ecological perspective. *Forensic Sci. Int.* 278, 388–395. doi: 10.1016/j.forsciint.2017.07.017
- Adserias-Garriga, J., Quijada, N. M., Hernandez, M., Lázaro, D. R., Steadman, D., and Garcia-Gil, J. (2017b). Dynamics of the oral microbiota as a tool to estimate time since death. *Mol. Oral Microbiol.* doi: 10.1111/omi.12191 [Epub ahead of print].
- Baker, G. C., Smith, J. J., and Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* 55, 541–555. doi: 10.1016/j.mimet.2003.08.009
- Boumba, V. A., Ziavrou, K. S., and Vougiouklakis, T. (2008). Biochemical pathways generating post-mortem volatile compounds co-detected during forensic ethanol analyses. *Forensic Sci. Int.* 174, 133–151. doi: 10.1016/j.forsciint.2007.03.018
- Burcham, Z. M., Hood, J. A., Pechal, J. L., Krausz, K. L., Bose, J. L., Schmidt, C. J., et al. (2016). Fluorescently labeled bacteria provide insight on post-mortem microbial transmigration. *Forensic Sci. Int.* 264, 63–69. doi: 10.1016/j.forsciint.2016.03.019
- Can, I., Javan, G. T., Pozhitkov, A. E., and Noble, P. A. (2014). Distinctive thanatomicrobiome signatures found in the blood and internal organs of humans. *J. Microbiol. Methods* 106, 1–7. doi: 10.1016/j.mimet.2014.07.026
- Clement, C., Hill, J. M., Dua, P., Culicchia, F., and Lukiw, W. J. (2016). Analysis of RNA from Alzheimer's disease post-mortem brain tissues. *Mol. Neurobiol.* 53, 1322–1328. doi: 10.1007/s12035-015-9105-6
- Corry, J. E. L. (1978). Possible sources of ethanol ante-mortem and postmortem—its relationship to biochemistry and microbiology of decomposition. *J. Appl. Bacteriol.* 44, 1–56. doi: 10.1111/j.1365-2672.1978.tb00776.x
- Damann, F. E., Williams, D. E., and Layton, A. C. (2015). Potential use of bacterial community succession in decaying human bone for estimating postmortem interval. *J. Forensic Sci.* 60, 844–850. doi: 10.1111/1556-4029.12744
- Edgar, R. C. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Edgar, R. C. (2004b). “MUSCLE: low-complexity multiple sequence alignment with T-coffee accuracy,” in *Proceedings of the Computational Systems Bioinformatics Conference, 2004: ISMB/ECCB*, Stanford, CA.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Gevers, W. (1975). Biochemical aspects of cell death. *Forensic Sci.* 6, 25–29. doi: 10.1016/0300-9432(75)90220-4
- Gill, J. R., and Landi, K. (2011). Putrefactive rigor: apparent rigor mortis due to gas distension. *Am. J. Forensic Med. Pathol.* 32, 242–244. doi: 10.1097/PAF.0b013e3181dd17b9
- Harrington, D. J. (1996). Bacterial collagenases and collagen-degrading enzymes and their potential role in human disease. *Infect. Immun.* 64, 1885–1891.
- Hyde, E. R., Haarmann, D. P., Lynne, A. M., Bucheli, S. R., and Petrosino, J. F. (2013). The living dead: bacterial community structure of a cadaver at the onset and end of the bloat stage of decomposition. *PLOS ONE* 8:e77733. doi: 10.1371/journal.pone.0077733
- Hyde, E. R., Haarmann, D. P., Petrosino, J. F., Lynne, A. M., and Bucheli, S. R. (2015). Initial insights into bacterial succession during human decomposition. *Int. J. Legal Med.* 129, 661–671. doi: 10.1007/s00414-014-1128-4
- Janaway, R., Percival, S., and Wilson, A. (2009). “Decomposition of human remains,” in *Microbiology and Aging*, ed. S. Percival (New York City, NY: Humana Press), 313–334.
- Javan, G. T., Finley, S. J., Abidin, Z., and Mulle, J. G. (2016a). The thanatomicrobiome: a missing piece of the microbial puzzle of death. *Front. Microbiol.* 7:225. doi: 10.3389/fmicb.2016.00225
- Javan, G. T., Finley, S. J., Can, I., Wilkinson, J. E., Hanson, J. D., and Tarone, A. M. (2016b). Human thanatomicrobiome succession and time since death. *Sci. Rep.* 6:29598. doi: 10.1038/srep29598
- Johnson, H. R., Trinidad, D. D., Guzman, S., Khan, Z., Parziale, J. V., DeBruyn, J. M., et al. (2016). A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PLOS ONE* 11:e0167370. doi: 10.1371/journal.pone.0167370
- Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6955–6959. doi: 10.1073/pnas.82.20.6955
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* 8:e61217. doi: 10.1371/journal.pone.0061217
- Metcalfe, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., Hyde, E. R., et al. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 351, 158–162. doi: 10.1126/science.aad2646
- Mondor, E. B., Tremblay, M. N., Tomberlin, J. K., Benbow, E. M., Tarone, A. M., and Crippen, T. L. (2012). The ecology of carrion decomposition. *Nat. Educ. Knowledge* 3, 21.
- Morosyuk, S. V., Lee, K., SantaLucia, J., and Cunningham, P. R. (2000). Structure and function of the conserved 690 hairpin in *Escherichia coli* 16 S ribosomal RNA: analysis of the stem nucleotides. *J. Mol. Biol.* 300, 113–126. doi: 10.1006/jmbi.2000.3852
- Morosyuk, S. V., SantaLucia, J., and Cunningham, P. R. (2001). Structure and function of the conserved 690 hairpin in *Escherichia coli* 16 S ribosomal RNA. III. Functional analysis of the 690 loop. *J. Mol. Biol.* 307, 213–228. doi: 10.1006/jmbi.2000.4432
- Morris, J. A., Harrison, L. M., and Partridge, S. M. (2006). Postmortem bacteriology: a re-evaluation. *J. Clin. Pathol.* 59, 1–9. doi: 10.1136/jcp.2005.028183
- Oksanen, J. (2011). *Multivariate Analysis of Ecological Communities in R: Vegan Tutorial. R Package Version*. Available at: <https://cran.r-project.org>
- Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., et al. (2009). The NIH human microbiome project. *Genome Res.* 19, 2317–2323. doi: 10.1101/gr.096651.109
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLOS ONE* 2015:e9490. doi: 10.1371/journal.pone.0009490
- Proskuryakov, S. Y., Konoplyannikov, A. G., and Gabai, V. L. (2003). Necrosis: a specific form of programmed cell death? *Exp. Cell Res.* 283, 1–16. doi: 10.1016/S0014-4827(02)00027-7
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., et al. (2000). Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* 102, 615–623. doi: 10.1016/S0092-8674(00)00084-2
- Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., et al. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827–834. doi: 10.1126/science.1117230
- Shah, V., Luxton, T. P., Walker, V. K., Brumfield, T., Yost, J., Shah, S., et al. (2016). Fate and impact of zero-valent copper nanoparticles on geographically-distinct soils. *Sci. Total Environ.* 573, 661–670. doi: 10.1016/j.scitotenv.2016.08.114
- Thomas, T. B., Finley, S. J., Wilkinson, J. E., Wescott, D. J., Gorski, A., and Javan, G. T. (2017). Postmortem microbial communities in burial soil layers of skeletonized humans. *J. Forensic Leg. Med.* 49, 43–49. doi: 10.1016/j.jflm.2017.05.009
- Tringe, S. G., and Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.* 11, 442–446. doi: 10.1016/j.mib.2008.09.011
- Vorobev, A. V., Baani, M., Doronina, N. V., Brady, A. L., Liesack, W., Dunfield, P. F., et al. (2011). *Methyloferula stellata* gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. *Int. J. Syst. Evol. Microbiol.* 61(Pt 10), 2456–2463. doi: 10.1099/ijso.0.028118-0

- Wan, M., Rosenberg, J. N., Faruq, J., Betenbaugh, M. J., and Xia, J. (2011). An improved colony PCR procedure for genetic screening of *Chlorella* and related microalgae. *Biotechnol. Lett.* 33, 1615–1619. doi: 10.1007/s10529-011-0596-6
- Wang, Y., and Qian, P. Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLOS ONE* 4:e7401. doi: 10.1371/journal.pone.0007401
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Willardsen, R. R., Busta, F. F., and Allen, C. E. (1979). Growth of *Clostridium perfringens* in three different beef media and fluid thioglycollate medium at static and constantly rising temperatures. *J. Food Prot.* 42, 144–148. doi: 10.4315/0362-028X-42.2.144
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M. Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., et al. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339. doi: 10.1038/35030006
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576
- Yang, B., Wang, Y., and Qian, P. Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 17:135. doi: 10.1186/s12859-016-0992-y
- Zhao, Y., Jaber, V., and Lukiw, W. J. (2017). Secretory products of the human GI tract Microbiome and their potential impact on Alzheimer's disease (AD): detection of lipopolysaccharide (LPS) in AD hippocampus. *Front. Cell. Infect. Microbiol.* 7:318. doi: 10.3389/fcimb.2017.00318
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2017 Javan, Finley, Smith, Miller and Wilkinson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership