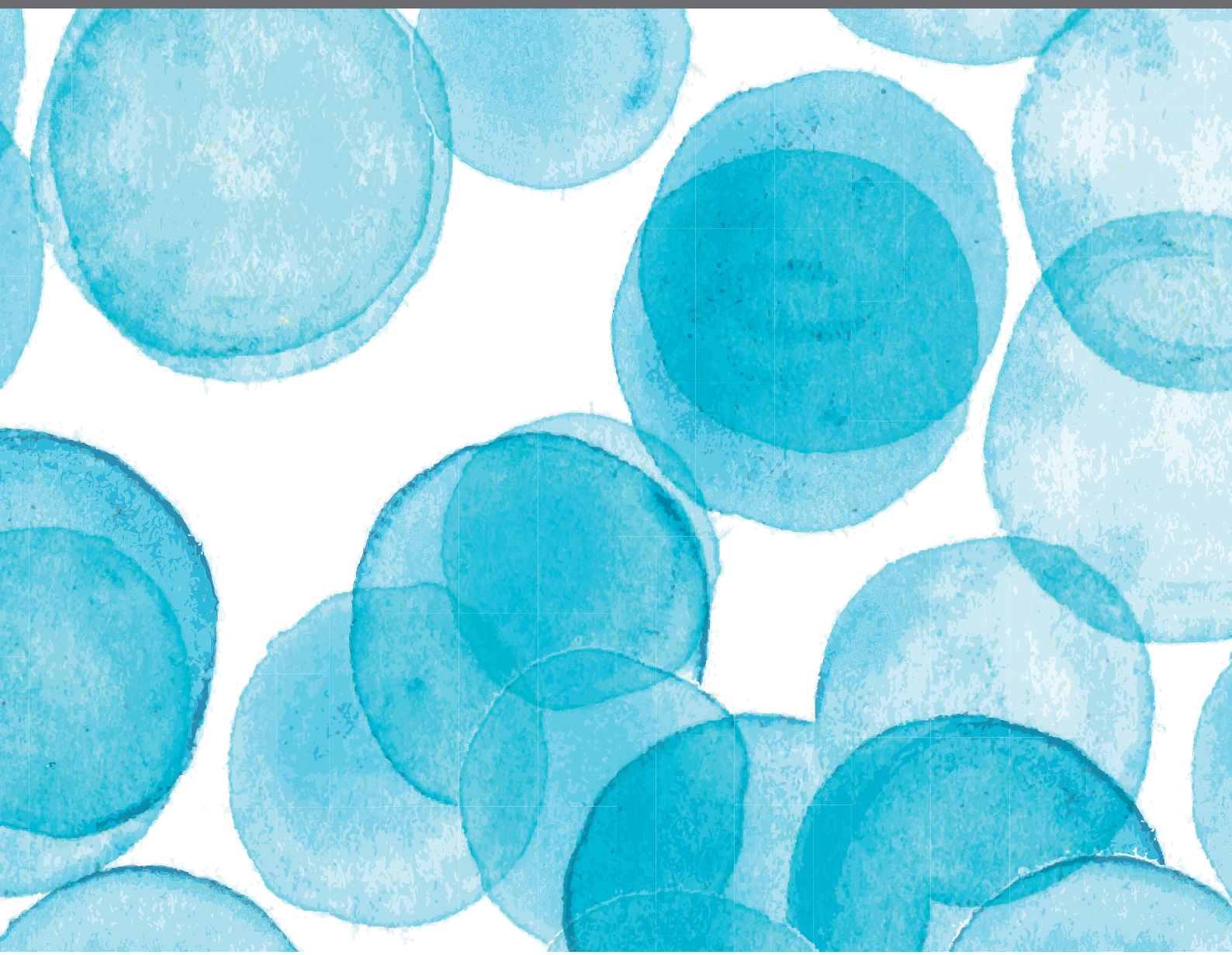
INSIGHTS IN EVOLUTIONARY AND GENOMIC MICROBIOLOGY: 2021

EDITED BY: Daniel Yero, Feng Gao and Baolei Jia
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-340-5
DOI 10.3389/978-2-88976-340-5

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: frontiersin.org/about/contact

INSIGHTS IN EVOLUTIONARY AND GENOMIC MICROBIOLOGY: 2021

Topic Editors:
Daniel Yero, Universidad Autónoma de Barcelona, Spain
Feng Gao, Tianjin University, China
Baolei Jia, Chung-Ang University, South Korea

Citation: Yero, D., Gao, F., Jia, B., eds. (2022). Insights in Evolutionary and Genomic Microbiology: 2021. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-340-5

Table of Contents

frontiers | Frontiers in Microbiology

EDITORIAL
published: 18 May 2022
doi: 10.3389/fmicb.2022.915593

Editorial: Insights in Evolutionary and Genomic Microbiology: 2021

Daniel Yero[1,2]*, *Baolei Jia*[3] *and Feng Gao*[4,5]

[1] Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, [2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, [3] Department of Life Science, Chung-Ang University, Seoul, South Korea, [4] Department of Physics, School of Science, Tianjin University, Tianjin, China, [5] Frontiers Science Center for Synthetic Biology, Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China

Keywords: next generation sequencing, long-read sequencing, comparative genomics, adaptive evolution, pan-genome

Editorial on the Research Topic

Insights in Evolutionary and Genomic Microbiology: 2021

Recent advances in DNA sequencing and bioinformatics have shaped and accelerated recent advances in all fields of microbiology (Gao, 2019; Kobras et al., 2021). Next-generation sequencing (NGS) together with large-scale genomic comparisons allows, among others, identifying the extent of genetic variability among related isolates and strains, and a more comprehensive analysis of the evolution of adaptation, antibiotic resistance, and pathogenicity factors in clinically relevant pathogens. In line with this, one study by Whaley et al. characterizes the capsule polysaccharide synthesis genes in 1,514 isolates of *Neisseria meningitidis* by whole-genome sequencing (WGS). This study shows that among meningococcal carriage isolates collected from student populations at three US universities, several mutations allow the bacteria to switch between an encapsulated and non-encapsulated state. The importance of this study lies in the fact that variations in capsule polysaccharide synthesis may potentially result in polysaccharide-based meningococcal vaccine escape. A second study by Manoharan-Basil et al. provides a current comprehensive overview of the role that horizontal gene transfer (HGT) plays in the evolution of antibiotic resistance in *Neisseria gonorrhoeae* by using WGS data comprising 20,047 isolates. This study focuses on the global spread and emergence of fluoroquinolone resistance in gonococci and highlights the significant role that HGT plays in transferring resistance from commensal *Neisseria* species to pathogenic *Neisseria*.

In a multinational and multicenter study by Topaz et al., authors present the phylogenetic structure and comparative genomics of 410 invasive *Haemophilus influenzae* serotype a (Hia) isolates. In addition to identifying genetic differences in virulence and antimicrobial resistance genes and describing the phylogenetic structure of Hia, a genome-wide association study was conducted in order to detect associations with clinical and epidemiological traits. They show that very little of the phenotypic variation was due to the genetic variability present in the studied population, at least for the variables analyzed. Another study by Yuan et al. presents a pan-genome analysis of *Laribacter hongkongensis*, a potential emerging pathogen, which leads to the study of the distribution of virulence and antimicrobial resistance genes among strains of different sources. Authors assess the influence of the source and lineage of strains on pathogenicity risk, suggesting that strains isolated from frogs may have a higher potential to become human pathogens.

OPEN ACCESS

Edited and reviewed by:
Ludmila Chistoserdova, University of Washington, United States

***Correspondence:**
Daniel Yero
daniel.yero@uab.cat

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

Received: *08 April 2022*
Accepted: *12 April 2022*
Published: *18 May 2022*

Citation:
Yero D, Jia B and Gao F (2022) Editorial: Insights in Evolutionary and Genomic Microbiology: 2021. Front. Microbiol. 13:915593. doi: 10.3389/fmicb.2022.915593

The increase in genomic information for some bacterial species has allowed molecular taxonomy and environmental adaptation to be studied in depth. For instance, the complete genome of *Acidiferrobacter thiooxydans* is presented for the first time in a work by Ma et al. (only a partial genome at NCBI at the time of this publication). Analysis of this sequence and comparative genomic studies reveal a unique phylogenetic position and a genomic plasticity that contributes to the environmental adaptation of these bacteria in extremely acidic niches with high metal concentrations. In another study by Kong et al., the complete genome of a new strain of *Rahnella victoriana* also reveals new insights into its molecular and genetic mechanism for promoting plant growth. Among the beneficial features, the new genome carries genetic information for indole-3-acetic acid production, volatile organic compound biosynthesis, nitrogen fixation, phosphate solubilization, siderophores, acetoin, 1-aminocyclopropane-1-carboxylate deaminase, and gamma-aminobutyric acid production.

In this Research Topic some studies use long-read WGS, a third-generation sequencing technology, to generate genomic data. Long-read techniques offer advantages over short-read sequencing because they are well suited to the detection of larger sequence changes such as structural variants and phased variations, and they have the ability to span repetitive regions (Amarasinghe et al., 2020). A study by Thibau et al. shows a conserved genome sequence among eight *Bartonella henselae* isolates and identifies a variable genomic island by using PacBio single molecule real-time (SMRT) sequencing. In a second study by Abdel-Glil et al., genome sequencing using the PacBio SMRT method was used to establish phylogenetic relatedness and the genome structure of *Yersinia ruckeri* isolates. Comparative genomics shows that *Y. ruckeri* has unique genomic regions probably related to the pathogenesis of enteric redmouth disease in fish.

Recent advances in the sequencing and analysis of bacterial genomes have also shed light on the adaptive evolution mechanisms. One study by Kurokawa et al. presents works on experimental evolution in *Escherichia coli* strains with different genome sizes with the aim to determine a connection between genome reduction with adaptation to environmental gradients and the contribution of mutations to bacterial fitness. The results reveal a quantitative relationship among genome reduction, adaptation, and niche expansion. A second study by Zhao et al. focuses on genomic island evolution as important adaptive traits in pathogenic or environmental bacteria. By using comparative genomics, a new composite island was identified and characterized in *Shewanella putrefaciens* carrying a functional HipAB toxin–antitoxin system. Finally, another study by Garzón et al. explored a novel *in silico* approach based on bacterial genomic data to propose a minimal translational machinery needed for protein synthesis. This information could contribute to the design of synthetic cells for biotechnological purposes.

AUTHOR CONTRIBUTIONS

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

FUNDING

FG is supported by the National Key Research and Development Program of China (Grant Number 2018YFA0903700) and the National Natural Science Foundation of China (Grant Numbers 21621004 and 31571358).

REFERENCES

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 30. doi: 10.1186/s13059-020-1935-5

Gao, F. (2019). Recent developments of software and database in microbial genomics and functional genomics. *Brief. Bioinform.* 20, 732–734. doi: 10.1093/bib/bby013

Kobras, C. M., Fenton, A. K., and Sheppard, S. K. (2021). Next-generation microbiology: from comparative genomics to gene function. *Genome Biol.* 22, 123. doi: 10.1186/s13059-021-02344-9

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fmicb.2021.782415

Phylogenetic Relatedness and Genome Structure of *Yersinia ruckeri* Revealed by Whole Genome Sequencing and a Comparative Analysis

Mostafa Y. Abdel-Glil[1], Uwe Fischer[2], Dieter Steinhagen[3], Una McCarthy[4], Heinrich Neubauer[1] and Lisa D. Sprague[1]**

[1] Friedrich-Loeffler-Institut, Institute of Bacterial Infections and Zoonoses (IBIZ), Jena, Germany, [2] Friedrich-Loeffler-Institut, Institute of Infectiology, Greifswald-Insel Riems, Germany, [3] Fish Disease Research Unit, Institute for Parasitology, University of Veterinary Medicine Hannover, Hanover, Germany, [4] Marine Laboratory, Marine Scotland, Aberdeen, United Kingdom

OPEN ACCESS

Edited by:
Daniel Yero,
Universidad Autónoma de Barcelona, Spain

Reviewed by:
Andrey P. Anisimov,
State Research Center for Applied Microbiology and Biotechnology, Russia
Ran Duan,
National Institute for Communicable Disease Control and Prevention (CDC), China
Scott Van Nguyen,
Public Health Laboratory Division, United States

***Correspondence:**
Lisa D. Sprague
lisa.sprague@fli.de
Mostafa Y. Abdel-Glil
mostafa.abdel-glil@fli.de

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

Received: 24 September 2021
Accepted: 25 October 2021
Published: 18 November 2021

Citation:
Abdel-Glil MY, Fischer U, Steinhagen D, McCarthy U, Neubauer H and Sprague LD (2021) Phylogenetic Relatedness and Genome Structure of Yersinia ruckeri Revealed by Whole Genome Sequencing and a Comparative Analysis. Front. Microbiol. 12:782415. doi: 10.3389/fmicb.2021.782415

Yersinia ruckeri is the causative agent of enteric redmouth disease (ERM), a serious infection that affects global aquaculture with high economic impact. The present study used whole genome sequences to perform a comparative analysis on 10 *Y. ruckeri* strains and to explore their genetic relatedness to other members of the genus. *Y. ruckeri*, *Yersinia entomophaga*, and *Yersinia nurmii* formed a species complex that constitutes the most basal lineage of the genus. The results showed that the taxonomy of *Y. ruckeri* strains is better defined by using a core genome alignment and phylogenetic analysis. The distribution of accessory genes in all *Yersinia* species revealed the presence of 303 distinctive genes in *Y. ruckeri*. Of these, 169 genes were distributed in 17 genomic islands potentially involved in the pathogenesis of ERM via (1) encoding virulence factors such as Afp18, Yrp1, phage proteins and (2) improving the metabolic capabilities by enhancing utilization and metabolism of iron, amino acids (specifically, arginine and histidine), and carbohydrates. The genome of *Y. ruckeri* is highly conserved regarding gene structure, gene layout and functional categorization of genes. It contains various components of mobile genetic elements but lacks the CRISPR-*Cas* system and possesses a stable set of virulence genes possibly playing a critical role in pathogenicity. Distinct virulence plasmids were exclusively restricted to a specific clonal group of *Y. ruckeri* (CG4), possibly indicating a selective advantage. Phylogenetic analysis of *Y. ruckeri* genomes revealed the co-presence of multiple genetically distant lineages of *Y. ruckeri* strains circulating in Germany. Our results also suggest a possible dissemination of a specific group of strains in the United States, Peru, Germany, and Denmark. In conclusion, this study provides new insights into the taxonomy and evolution of *Y. ruckeri* and contributes to a better understanding of the pathogenicity of ERM in aquaculture. The genomic analysis presented here offers a framework for the development of more efficient control strategies for this pathogen.

Keywords: *Yersinia ruckeri*, genome, taxonomy, evolution, fish, enteric redmouth disease, phylogeny, cgMLST and SNPs

INTRODUCTION

The genus *Yersinia* currently comprises 26 species with different ecological habitats and pathogenicity. *Yersinia ruckeri* is a Gram negative, rod shaped, facultative intracellular enterobacterium and the causative agent of enteric redmouth disease (ERM), a serious septicaemic bacterial disease of salmonids. Fish in all stages of development are susceptible, resulting in a high degree of mortality (Guijarro et al., 2018). The clinical picture is characterized by hemorrhages in the skin and mucosa, as well as exophthalmia, darkening of the skin, inflammation of the lower intestine and splenomegaly (Kumar et al., 2015). Phenotypically, strains can be distinguished according to their biotypes (biotype 1 and biotype 2), serotypes (O1a, b; O2 a, b, c; O3; and O4) as well as their outer membrane protein (OMP) profiles (Romalde and Toranzo, 1993; Kumar et al., 2015; Wrobel et al., 2020). OMPs contribute significantly to the composition of the outer membrane of Gram-negative bacteria and play a major role in virulence. Studies on the composition of the outer membrane proteome of *Y. ruckeri* isolates from rainbow trout and Atlantic salmon revealed a high degree of variation between strains (Ormsby et al., 2019; Ormsby and Davies, 2021). Several pathogenicity factors have been implicated in the virulence of *Y. ruckeri* strains, including toxins, secretion systems, iron and amino acid utilization systems, lipopolysaccharides, flagella and plasmids. The current knowledge on virulence factors and infection pathways of *Y. ruckeri* is outlined in recently published reviews (Guijarro et al., 2018; Wrobel et al., 2019).

Yersinia ruckeri poses a serious threat to the global aquaculture industry and infection can result in significant economic losses. Disease management is currently based on vaccination and antibiotic treatment. The first commercial fish vaccine for ERM was a bacterin prepared from formalin inactivated whole cells of *Y. ruckeri* and licensed in 1976. However, outbreaks are being increasingly reported throughout the world (Kumar et al., 2015; Wrobel et al., 2019). This development of "vaccine breakthrough" strains necessitates the design of new and more effective vaccines based on data obtained from comparative genome studies. Moreover, these studies can elucidate how fish pathogens spread and evolve with aquaculture, help to create management strategies to improve biosecurity and ultimately significantly reduce financial losses (Ormsby and Davies, 2021; Yang et al., 2021).

Genome comparison studies enable detailed insights into understanding pathogenicity and evolution of bacteria and help to clarify transmission routes as well as the geographic spread of pathogens. In this study, (1) we describe the genomic features of ten circularized genomes of *Y. ruckeri* strains of mostly clinical origin; (2) explore the genetic relatedness of *Y. ruckeri* within the genus *Yersinia* by comparing our data with the genomic data of 90 representative strains of all validly published *Yersinia* species; (3) describe the comparative phylogenetic analysis and *in silico* virulence profiling of the ten sequenced strains with published genomes of 67 *Y. ruckeri* strains.

RESULTS

Yersinia ruckeri Strains and Genome Sequencing

Ten *Y. ruckeri* strains of predominantly clinical origin, i.e., diseased fish were characterized by traditional bacteriological methods and sequenced using Pacific Bioscience and Illumina MiSeq platforms. These included seven strains from Germany isolated between 2005 and 2011, two strains isolated in Scotland in 1999 and 2007, and the *Y. ruckeri* type strain (DSM18506/ATCC 29473; here: 16Y0180) (**Tables 1**, **2**).

Genome sequencing using the Pacific Bioscience SMRT® method produced on average 80,586 reads (70592–92078) and 900,928,849 bases (632,133,554–1.276,956,896) per strain with a mean sequencing depth of 178-fold (129 to 248-fold). The mean read length was 9,065.3 bp (1007–14225 bp) and the mean N50 value of the reads was 14929.9 bp (10855–19303 bp) (**Supplementary Table 1**). Sequencing using Illumina MiSeq generated on average 1,371,836 paired-end reads (664,572–2.680,262) and 364,034,507 bases (180,017,799–686,936,084) accounting for a mean theoretical sequencing depth of 98.4-fold (48.6 to 185.6-fold) based on the length of the reference *Y. ruckeri* genome, Big Creek 74 (accession GCF000964565.1) (**Supplementary Table 1**).

Overview of Features of Circularized *Y. ruckeri* Genomes

Genome Description

Sequenced genomes were assembled *de novo* and the resulting circularized genomes were iteratively polished with Illumina reads to obtain finished genomes with high fidelity (see section "Materials and Methods"). The complete genome of *Y. ruckeri* revealed a single circular chromosome with several distinct features (**Supplementary Figure 1**). The structure of the chromosomes was characterized by a high level of conservation between the strains with regard to genome length, GC content, RNA genes and coding capacity. The size of the chromosome ranged between 3.67 and 3.81 mega bases (Mb) with an average GC content of 47.5% (**Supplementary Table 1**). The density of coding DNA sequences (CDS) averaged 84.2% (84.02–84.46%) with total annotated CDS features averaging 3374 (3267–3460) per genome and average gene size of 939 bases (90–12,159 bp). The chromosome of *Y. ruckeri* contained 81–84 tRNA genes and seven ribosomal RNA operons comprising 16S, 23S, and 5S rRNA genes (**Supplementary Table 1**). The genome of the *Y. ruckeri* strains exhibited a comparatively high number of pseudogene candidates, which comprised on average 5.7% (4.4–7.1%) of the total number of coding sequences (**Supplementary Table 1**). Of the predicted pseudogenes, 40–50% of the pseudogenes represented fragmented CDS while 17–24% represented truncated genes.

Gene Layout

A comparison of the gene layout in the chromosome of the ten strains revealed a collinear genome with a limited frequency of

TABLE 1 | List of *Yersinia ruckeri* strains sequenced in this study.

Strain number	Alias	Host	Isolation Year	Origin
16Y0180	Type strain (DSM18506/ATCC 29473)	Rainbow trout	–	Idaho, United States
17Y0153	G1S1	Rainbow trout	2008	North Rhine-Westphalia, Germany
17Y0155	LT13-1/0811	Rainbow trout	2011	North Rhine-Westphalia, Germany
17Y0157	178-1/05	Brown trout	2005	Lower Saxony, Germany
17Y0159	285-1/05	Pike	2005	Lower Saxony, Germany
17Y0161	111-1/05	–	2005	Lower Saxony, Germany
17Y0163	1521/11	Rainbow trout	2011	Hesse, Germany
17Y0189	KP6	Rainbow trout	2011	North Rhine-Westphalia, Germany
17Y0412	MT2209	Atlantic salmon	1999	Highlands, Scotland, United Kingdom
17Y0414	MT3187	Atlantic salmon	2007	Western Isles, Scotland, United Kingdom

TABLE 2 | Phenotypic characterization of the ten *Yersinia ruckeri* strains.

Strain number	Biochemical identification	CO	Catalase	Gram staining	Morphology	Motility	API 20E	Glucose/Gas	Tween 80	Tween 20	Xylose	Nitrate	Gelatinase	Citrate	VP	MR	Sorbitol	OF-O/F	N2
16Y0180	*Y. ruckeri*	–	+	–	rods	+	1105100	+	nd	nd	–	nd	–	–	+	nd	–	nd	nd
17Y0153	*Y. ruckeri*	–	+	–	short rods	–	530710017	+/–	–	–	–	+	+	+	+	+	–	+/+	nd
17Y0155	*Y. ruckeri*	–	+	–	short rods	+	530710057	+/–	+	+	–	+	+	–	+	+	–	+/+	nd
17Y0157	*Y. ruckeri*	–	+	–	short rods	+	530750057	+/–	+	+	–	–	+	+	+	–	+	+/+	+
17Y0159	*Y. ruckeri*	–	+	–	rods	+	530570057	+/–	+	+	–	+	–	+	+	–	+	nd	nd
17Y0161	*Y. ruckeri*	–	+	–	short rods	+	530750047	+/–	+	+	–	+	+	+	+	–	+	+/+	nd
17Y0163	*Y. ruckeri*	–	+	–	short rods	+	530750057	+/–	+	+	–	+	+	+	+	–	+	+/+	nd
17Y0189	*Y. ruckeri*	–	+	–	rods	–	5107100	+	nd	nd	–	nd	+	–	+	nd	–	nd	nd
17Y0412	*Y. ruckeri*	–	+	–	rods	+	5107100	+	nd	nd	–	nd	+	–	+	nd	–	nd	nd
17Y0414	*Y. ruckeri*	–	+	–	rods	+	5107100	+	nd	nd	–	nd	+	–	+	nd	–	nd	nd

CO: cytochrome oxidase reaction; VP: Voges-Proskauer test; MR: methyl red test; OF-F/O: Oxidation-Fermentation; nd: not determined.

gene rearrangements in the majority of strains. In two strains, 17Y0161 and 17Y0163, a large segment of the chromosome (∼3 Mb) was symmetrically inverted relative to the replication fork (**Figure 1**) and was flanked by rRNA operons.

Functional Annotation of Genes

Functional annotation of the chromosomal genes using the Clusters of Orthologous Groups (COGs) classification scheme showed no significant variation between the distribution patterns of the COG groups among the genomes (standard deviation per COG group between genomes ranged from 0.8 to 12 genes; **Supplementary Table 2**). More than 90% of the genes were assigned to a specific COG group including the 18% of genes belonging to the COG category "S; function unknown." Of the genes annotated with known functions, the majority (45–47%) belonged to the "metabolism" group followed by the "cellular processes and signaling" (29–30%) and the "information storage and processing" (23–24%) groups (**Supplementary Table 2**). Some COG groups occurred with higher frequency such as the "amino acid transport and metabolism; group E" (20%), the "transcription, cell wall/membrane/envelope biogenesis; group M" (17%) and the "intracellular trafficking, secretion, and vesicular transport; group U" (15–17%). Contrariwise, only a few genes (∼2%) were assigned to the COG group "defense mechanisms" (**Supplementary Table 2**).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and Mobile Genetic Elements

No CRISPR-*Cas* regions were detected in any of the sequenced genomes. Therefore, the mobile element structures in the genomes i.e., plasmids, phages and insertion sequences were analyzed in depth. Seven of the ten genomes contained plasmids; six genomes contained a single plasmid and one genome three plasmids (**Supplementary Table 1**). By applying the Hadamard matrix, which combines values of percentage identity and alignment length between plasmid genome pairs, the nine plasmids could be allocated to six plasmid types. A 103 kb plasmid was present in four genomes while the other three genomes carried distinct plasmids of 23.8, 28.7, 59.4, 83.7, and 85.1 kb, respectively. The number of coding sequences varied between 25 and 101 CDS per plasmid. Most of the plasmid genes encoded hypothetical proteins. The COG annotation revealed that the function of 30% of the well-characterized plasmid genes was related to "replication, recombination and repair." Functions related to

metabolism were poorly represented in the plasmid genes (**Supplementary Table 3**).

Integrated and complete phages were predicted in the chromosome of nine strains based on PHASTER classification with a phage completeness score between 100 and 150 (Arndt et al., 2016); six strains were predicted to contain two phages, two strains to contain three phages and one strain to contain four phages (**Table 3**). Nine phage types were predicted in total. Each analyzed *Y. ruckeri* strain displayed a unique combination of prophages (**Table 3**). However, the phage "Salmon 118970 sal3" (NC_031940) from the *Myoviridae* family was present in eight strains (**Table 3**). Salmon 118970 sal3 has been reported in different bacterial species such as *Salmonella* Typhimurium and *Escherichia coli*, and was found to exhibit lytic activity against *Salmonella* strains (Paradiso et al., 2016). The GC content of the predicted phages was generally different to that of the *Y. ruckeri* genome, ranging between 45.1 and 49.8%.

TABLE 3 | Summary of mobile genetic elements in the complete genomes of the ten *Yersinia ruckeri* strains.

	Insertion sequences*										Intact phages in the chromosome (number, GC content%)**									
Strain number	IS1	IS200/IS605	IS256	IS3	IS481	IS5	ISL3	new	Total number	Percentage	GF_2 NC_026611	SfI NC_027339	SfV NC_003444	ENT90 NC_019932	500465_1 NC_049342	SuMu NC_019455	118970_sal3 NC_031940	SEN34 NC_028699	PY54 NC_005069	Length of predicted phages
16Y0180		3	1	21					25	0.94			1 (49.8%)		1 (46.2%)					33.4- and 35.4-kb
17Y0153		3	1	18				1	23	0.84			1 (49.8%)				1 (47.1%)			33.3- and 66.5-kb
17Y0155		3		18					21	0.79		1 (49.8%)					1 (47.8%)			33.4- and 51-kb
17Y0157	1	3	1	11					16	0.56										
17Y0159		1	5	12	1		5		24	0.85					1 (45.1%)	1 (49.2%)	1 (47.2%)			53. 5-, 43. 4-, and 38.9-kb
17Y0161		2		9		1			12	0.46				1 (49.2%)			1 (47.8%)			47.9- and 46.4-kb
17Y0163		3		6					9	0.35							1 (48.8%)	1 (48.7%)		45.9- and 30.8-kb
17Y0189		3		18					21	0.79							2 (47.1 and 49.8%)			51- and 33.3-kb
17Y0412		3	1	14	1				19	0.7	1 (47.3%)					1 (48.9%)	1 (48.5%)		1 (48.5%)	35. 6-, 41. 9-, 44. 9-, and 50.1-kb
17Y0414		3	1	14					18	0.64	1 (47.4%)						1 (48.5%)		1 (48.5%)	34. 1-, 44. 9-, and 50.1-kb

**Only complete insertion sequences are reported. **PHASTER completeness score 100–150.*

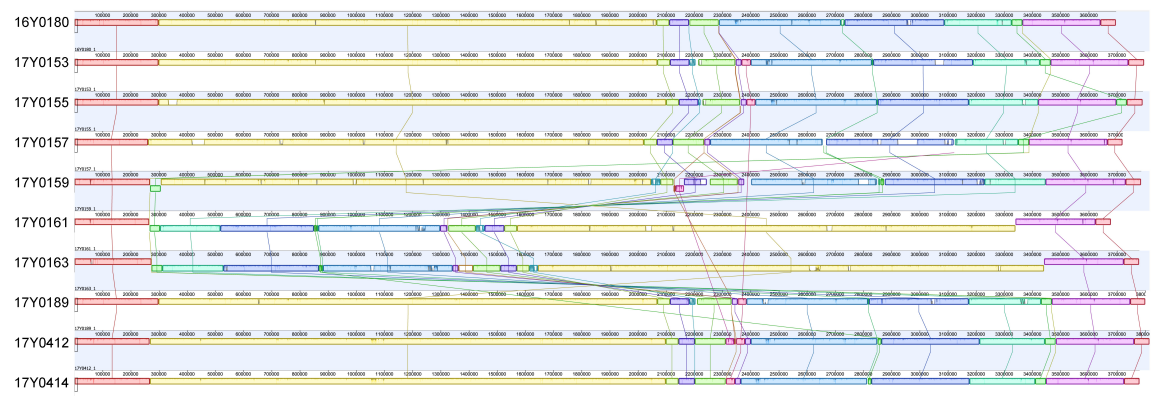


FIGURE 1 | Multiple genome alignment with progressiveMauve for the 10 *Yersinia ruckeri* strains relative to the type strain (16Y0180; DSM 18506); Syntenic homologous regions between strains are shown as colored blocks with a similarity profile corresponding to the average conservation of sequences within the blocks. Homologous blocks are displayed in matching colors and connected by vertical lines. Blocks placed below the center line are in inverse orientation. Regions outside the blocks represent unique DNA regions in the genomes. The scale is in base pairs.

Nine to twenty-five insertion sequences (IS) between 13 and 34 kb in size were predicted for each chromosome (**Table 3**), representing a variety of IS families, i.e., IS1, IS200/IS605, IS256, IS3, IS481, IS5, and ISL3. Most of the observed IS elements belonged to the IS3 family accounting for 50 to 85% of the total number of insertion sequences per genome (**Table 3**).

Comparative Analysis of *Y. ruckeri* to the Genus *Yersinia*

In order to explore the genetic relationship of the *Y. ruckeri* strains with the genus *Yersinia*, we additionally downloaded the genome sequence data of 90 strains representing the 26 currently validated *Yersinia* species (Savin et al., 2019; Le Guern et al., 2020; Nguyen et al., 2021). Reference genomes in the RefSeq database and the genomes of the species type strains were prioritized during data retrieval from NCBI. The genomic features of the downloaded WGS data are summarized in **Supplementary Table 4**.

16S rRNA Gene Analysis

The rRNA genes were extracted from the ten sequenced *Y. ruckeri* genomes. The rRNA genes were present in seven almost identical copies across the genome with a sequence similarity of 99–100%. Phylogenetic analyses of the 16S rRNA region of the 10 *Y. ruckeri* strains and the representative *Yersinia* species (n = 88) showed a clear separation of *Y. ruckeri* from the other *Yersinia* species of the genus, with all *Y. ruckeri* strains clustering in a monophyletic lineage (**Supplementary Figure 2**). The 16S rRNA genes of *Y. ruckeri* were highly conserved and showed more than 99.1% sequence identity (**Supplementary Table 5**). Based on 16S rRNA gene analysis, the closest related species to *Y. ruckeri* is *Yersinia kristensenii* with an average nucleotide similarity of 99% (**Supplementary Table 5**).

Genomic Average Nucleotide Identity

The genomic average nucleotide identity (gANI) was estimated by performing a pairwise alignment of genome stretches with MUMmer (Pritchard et al., 2016). The ANI values between the *Y. ruckeri* strains averaged 99% which exceeded the 95–96% ANI cut-off generally used to define species boundaries (Richter and Rossello-Mora, 2009; **Supplementary Figure 3A**). In contrast, ANI values between the *Y. ruckeri* strains and the different *Yersinia* species averaged around 84% nucleotide similarity (**Supplementary Figure 3B**). The alignment fraction used to estimate nucleotide identity, ranged between 15 and 100% (**Supplementary Figure 3C**). Sequence alignment fractions encompassing more than 50% of the genome grouped the 26 *Yersinia* species into three distinct species complexes (**Supplementary Figure 3C** and **Figure 2**). *Y. ruckeri*, *Yersinia entomophaga*, and *Yersinia nurmii* form species complex 1 while species complex 2 contains *Yersinia similis*, *Yersinia pseudotuberculosis*, *Yersinia pestis*, and *Yersinia wautersii*. Species complex 3 includes all other *Yersinia* species (n = 19) (**Figure 2**). The independent ANI algorithms FastANI and Microbial Species Identifier (MiSI) corroborated these results (**Supplementary Table 6**).

Core-Genome-Based Phylogeny

The soft-core genome encompasses 2,002 genes (2,046,925 bp) present in over 95% of the *Yersinia* genomes. The consensus SNP positions comprise 25% (n = 512,502) of the concatenated gene alignments, with 96% (n = 491,746) of them parsimony-informative. The phylogenetic analysis based on the core nucleotide alignment was performed by inferring a mid-point rooted maximum-likelihood tree (**Figure 2**). The branching pattern of the phylogeny was consistent with the ANI results, showing that the species of this genus grouped into three monophyletic lineages corresponding to the three species-complexes (**Figure 2**), with *Y. ruckeri* belonging to species complex 1. This species complex constitutes the most basal clade that roots the *Yersinia* genus (**Figure 2**).

Core Genome Multilocus Sequence Typing Allele Typing

Allele typing of the core genome was applied to our data (February 2021) using the recently published *Yersinia* core genome multilocus sequence typing (cgMLST) scheme (Savin et al., 2019). Two genomes (17Y0153 and 17Y0155) were assigned to the core-genome sequence type (cgST) 435. The results also showed that 13 of the 500 cgMLST loci were consistently absent in all ten *Y. ruckeri* genomes. A minimum spanning tree based on the allelic profiles of the 100 analyzed *Yersinia* strains revealed long allelic distances for the 500 loci among the genomes (**Figure 3**). Distance thresholds defining a *Yersinia* species on genus-based cgMLST were found to vary greatly between the species as previously described (Savin et al., 2019). Based on allele typing of the cgMLST scheme, the *Y. ruckeri* genomes grouped together, with pairwise allelic differences varying between 2 and 434. However, the allelic distances to the neighboring species were 476 to *Y. entomophaga*, 479 to *Yersinia rohdei*, 480 to *Yersinia aldovae* and 481 to *Y. pestis*.

Distribution of Accessory Genes

Next, we sought to identify the genes that are significantly associated with the *Y. ruckeri* species. We calculated a pangenome of 25,498 genes for the 100 *Yersinia* strains, of these 23,496 corresponded to accessory genes. A pan-genome-wide association study (GWAS) approach identified 34 genes significantly underrepresented in *Y. ruckeri* compared to the other *Yersinia* species (100% sensitivity and specificity, see section "Materials and Methods") (**Supplementary Table 7**). The missing genes coded predominantly for putative transporter proteins. Contrariwise, 303 genes had a statistically significant association with *Y. ruckeri* and their uniqueness was confirmed using BLASTN at 90% identity and 80% coverage (**Supplementary Table 7**). The COG annotation assigned the majority of genes to metabolism (n = 103), cell wall/membrane/envelope biogenesis (n = 30), intracellular trafficking, secretion and vesicular transport (n = 26), transcription (n = 20), and cell motility (n = 11) (**Supplementary Table 8**). Ninety-three genes were predicted to encode cytoplasmic proteins, 113 to encode cytoplasmic membrane proteins, 4 to encode extracellular proteins.

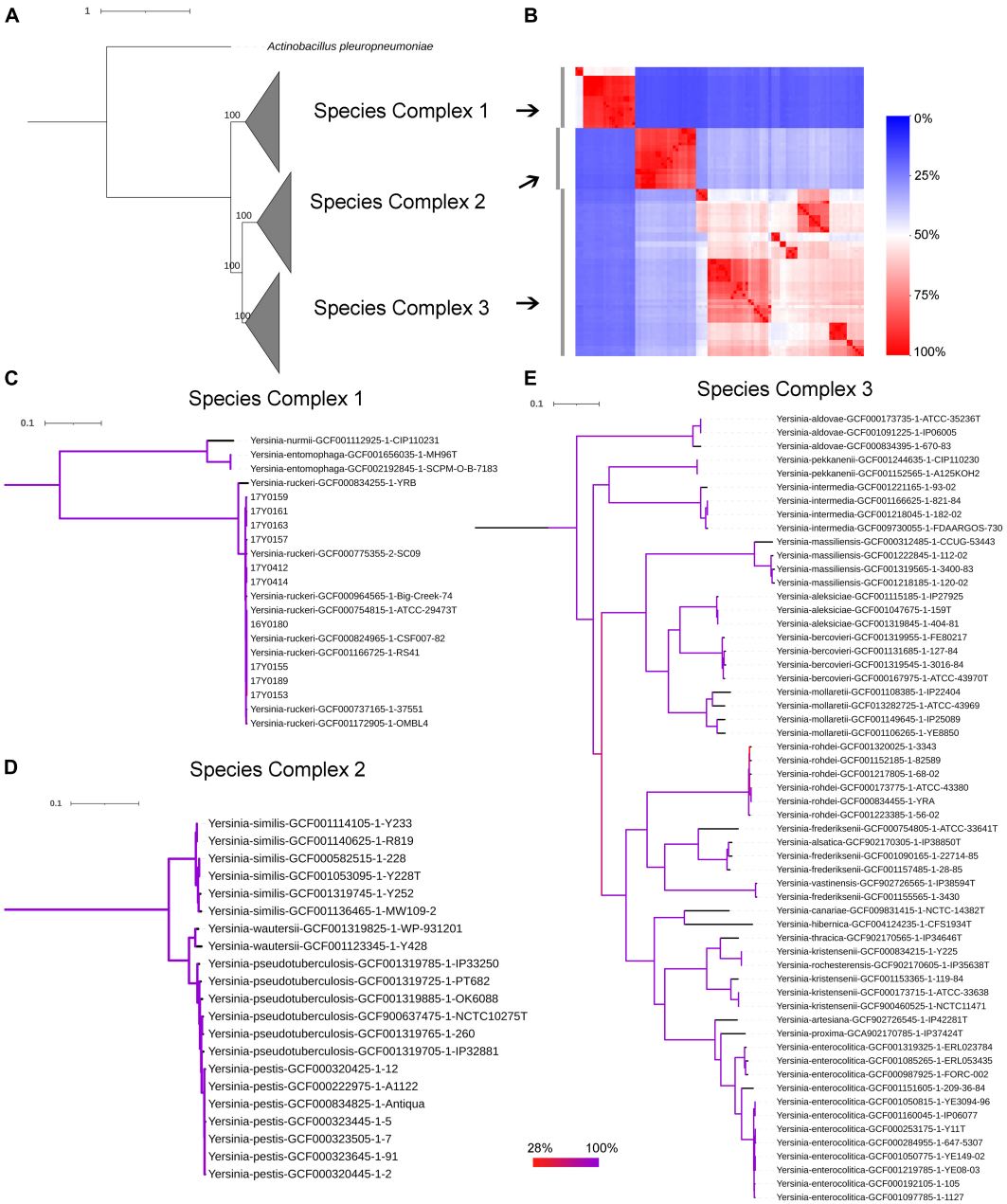


FIGURE 2 | Phylogenetic relationships within the genus *Yersinia*. **(A)** Maximum likelihood (ML) phylogenetic analysis based on the core protein alignment with the inclusion of strain *Actinobacillus pleuropneumoniae* as an outgroup strain (strain S4074, accession NZ_CP030753.1). **(B)** A heat map showing the alignment coverage for nucleotide identity calculation. **(C–E)** ML phylogenetic analysis based on the core nucleotide alignment of 2,002 genes (2,046,925 bp) present in more than 95% of the investigated 100 strains. The panels **(C–E)** correspond to species complex 1, 2, and 3, respectively. Branch coloration refers to the bootstrap support based on the analysis of 100 resampled trees.

Of the 303 *Y. ruckeri*-specific genes, 169 were arranged into 17 gene clusters/regions (1.5–49.6 kb). The largest genomic region (49 kb) included genes predominantly involved in protein utilization e.g., the *Y. ruckeri* protease 1 (Yrp1)-encoding operon that comprises *yrp1*, encoding for serralysin metalloprotease (EC 3.4.24.40), *inh* for a protease inhibitor and *yrpD*, *yrpE*, and *yrpF* for type I ABC protease exporters. In addition, genes encoding for histidine utilization proteins were predicted including EC 3.5.1.68 (N-formylglutamate deformylase), EC 3.5.2.7 (Imidazolonepropionase), histidine utilization repressor, EC 3.5.3.13 (Formiminoglutamic iminohydrolase), and HutD. Furthermore, an operon encompassing five successive genes putatively involved in the arginine succinyltransferase (AST) pathway for arginine metabolism, were predicted including EC 2.6.1.81 (succinylornithine transaminase), EC 2.3.1.109 (arginine N-succinyltransferase), EC 1.2.1.71

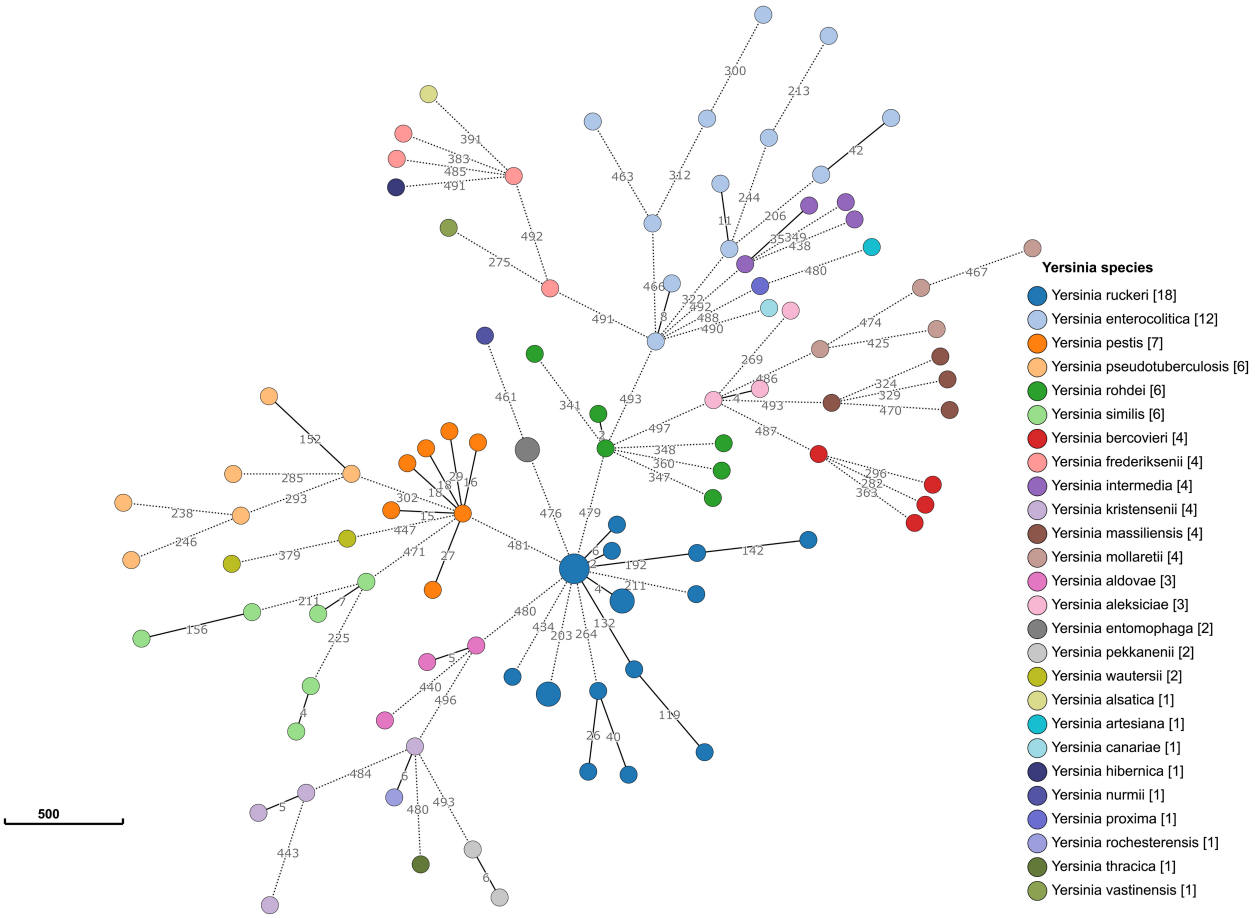


FIGURE 3 | Minimum spanning tree of the genus-based cgMLST for *Yersinia* strains. The dataset includes 90 representative genomes of 26 different *Yersinia* species as well as the ten strains sequenced in the study. Allelic profiles were calculated using the Genome Comparator function of BIGSdb. Nodes (circles) were colored according to species. Each node corresponds to a single core genome ST. The size of the node corresponds to the number of strains. The number of different alleles is shown on the branches (lines) between the nodes. Branches represented by dotted lines indicate more than 200 allelic differences between the strains.

(succinylglutamate-semialdehyde dehydrogenase), EC 3.5.3.23 (N-succinylarginine dihydrolase), and EC 3.5.1.96 (succinylglutamate desuccinylase). A gene encoding an OMP complex (OMPC) was found between the arginine and histidine utilization operons. No mobile elements were detected in this large genomic region.

The virulence gene of the antifeeding prophage 18 (*Afp18*) was detected in the second largest genomic region (30 kb). Further downstream, the transposase gene *IS285* bordered *Afp18*. Three additional genes with similar "left" orientation bordered by a transposase gene, ISKpn20 of the IS3 family, were found upstream of *Afp18* suggesting a putative operon structure. In the type strain, the *Y. ruckeri* invasin-like molecule (yrIlm) was located upstream of this genomic region and flanked by the ISEc39 transposase gene of the IS256 family. Phage proteins representing an incomplete phage of 7.7 kb were also predicted. This incomplete phage was similar to the Bacill_BCD7 phage (NC_019515).

The third genomic island (30 kb) included several genes encoding for a phosphotransferase system (PTS) involved in the uptake of carbohydrates (transport of N-acetylgalactosamine) by catalyzing the phosphorylation of sugar substrates. Three genes encoding sialidases flanked this region; two genes for N-acetylneuraminate epimerase (EC 5.1.3.24) and one gene for N-acetylneuraminic acid outer membrane channel protein NanC. An ISSpr1 transposase of the IS3 family was located downstream of the sialidase genes. A PTS system was additionally predicted in a further 25 kb genomic island. The latter harbored a cold-shock protein belonging to the CSP family.

A 21 kb gene cluster was detected comprising 15 genes, primarily involved in iron transport including a catechol siderophore system known as ruckerbactin, an important pathogenicity factor for *Y. ruckeri* infection in fish. A further 20 kb-genomic region was flanked by TnBth2 transposase of the Tn3 family. This region included three homolog genes to the SsrAB two-component regulatory system present within the *Salmonella* pathogenicity island (SPI-2). SsrAB is essential for survival and replication within host immune cells. Two type II secretion systems previously described in the genome of *Y. ruckeri* strain SC09 were detected in association with

two genomic regions; one thereof was flanked by an ISSpr1 transposase gene. However, the two T2SS systems were not exclusively present in *Y. ruckeri* genomes. **Supplementary Table 9** lists the composition and predicted annotation function of the identified *Y. ruckeri*-specific gene clusters.

Comparative Analysis of *Y. ruckeri*

Y. ruckeri Pangenome

To explore the diversity of the species *Y. ruckeri*, the sequence data of the ten *Y. ruckeri* strains sequenced in this study were combined with 67 *Y. ruckeri* genomes from prior studies (**Supplementary Table 10**). A total of 263,461 protein-coding sequences was present in the 77 strains. The pangenome analysis clustered the coding sequences into 5,655 gene clusters corresponding to the entirety of non-redundant genes. The core genome comprised 52% of the pangenome genes; 2,955 genes (2,892,182 bp) were present in more than 95% of the strains and 2,752 genes were present in all strains. A divergence of 2.4% was observed in the core genome for *Y. ruckeri*. SNP sites in the core orthologous genes covered 68.5 kb. The accessory genome comprised 2,700 genes distributed in less than 95% of the strains. Of these, 794 genes were strain-specific i.e., exclusively present in one strain. The pangenome was in its open form as previously demonstrated (Barnes et al., 2016). The trajectory pattern of the curve depicting the pangenome increased by adding more genomes. This openness was confirmed (Bpan = 0.25) by applying the regression models proposed by Tettelin et al. (2008). A limited reduction in the size of the core genome was observed. The core genome accounted for 85.3% of the coding sequences in the *Y. ruckeri* genome (n. CDS = 3,225, Big Creek 74, NZ_CP011078.1), while the core versus pangenome ratio was 48.6%. **Supplementary Figure 4** shows the trending pattern of the pangenome increase versus core genome decrease with sequential addition of genomes. The curve representing the new gene detection did not converge to zero, with new genes continuously contributing to the expansion of the pangenome.

Core Genome Phylogeny

Parsnp alignment followed by Gubbins-based filtration of loci with elevated densities of base substitutions identified 60,309 core SNP sites that represented putative point mutations outside recombination regions. One to 157 putative recombination blocks were identified containing five to 9,514 SNPs. The ratio of SNPs within recombination to the SNPs outside recombination (r/m) was 0,128, indicating a low effect of homologous recombination on the population diversity of *Y. ruckeri*. The core genome SNPs included a remarkable accumulation of non-synonymous variations (n = 58,972 SNPs) changing the amino acid sequence and possibly altering protein function. 17,336 SNPs represented silent (synonymous) gene changes, 13,768 SNPs were present in the intergenic regions while 733 SNPs were in other structures such as RNA.

The ML phylogenetic tree generated using RAxML based on non-recombinant SNPs is shown in **Figure 4** along with a heat map of pairwise SNP differences between all genomes. The genome of strain YRB was the most distant in the phylogeny. The other strains were grouped into three phylogenetic lineages. Lineage 1 previously described by Barnes et al. (2016) contained strains from Norway, Australia, China, and New Zealand. This lineage included two additional strains from Scotland and one strain from Germany. Three further strains from Germany were placed in lineage 2. Lineage 3 encompassed four strains, three from Germany and one of unknown origin. The German strains exhibited pairwise SNP distances averaging 4,037 SNPs, while the two strains from Scotland displayed only four SNPs.

Grouping the strains on fewer than 50 recombination-purged pairwise SNPs revealed the presence of six clonal groups (CG) with specific geographic associations. CG1 contained strains from Scotland, CG2 from Australia, CG3 from New Zealand, CG5 from both Australia and New Zealand, and CG6 from Australia, with the exception of one strain from Chile. CG4 included strains from the United States, Germany, Peru, and Denmark.

Virulence Genes in *Y. ruckeri*

Comparative analysis of virulence genes revealed the presence of a stable set of virulence factors in all *Y. ruckeri* strains (**Figure 5**). These include toxin genes for the extracellular Yrp1, *Y. ruckeri* peptidases (YrpAB) and *Y. ruckeri* pore forming toxin (yhlBA) as well as the Afp18. Additionally, the *ZnuABC* operon encoding a zinc binding protein and the BarA-uvrY regulator two-component systems were detected in all strains, as well as the cdsAB operon important for cystine transport and utilization, the flhDC operon involved in the regulation of the flagellar secretion system, and *lpxD* involved in lipid A biosynthesis. All these factors have previously been shown to be linked to the pathogenicity of *Y. ruckeri* (Wrobel et al., 2019). The recently described invasion gene *Y. ruckeri* invasin (*yrInv*) (Wrobel et al., 2018a) was detected in 20 strains.

The secretion systems described in strain SC09 included a type III secretion system (T3SS). Two type II secretion systems (T2SS; Yst1 and Yst2) were detected in all strains with the exception of strain OMBL4.

Strains belonging to CG4 as well as strain NHV3758 carried additional virulence factors, the type IV pili system encoded by the *pil* operon and the type IV secretion system encoded by the *tra* operon. As previously reported, these elements are located on a large (80–103 kb) potential virulence plasmid (Wrobel et al., 2018b).

DISCUSSION

Exploring the Evolutionary Relatedness of *Y. ruckeri* to the Genus *Yersinia*

The present study performed a comparative analysis on the genomes of the fish pathogen *Y. ruckeri* and explored its phylogenetic relatedness to the further members of the genus *Yersinia*. The results showed that the taxonomic assignment of *Y. ruckeri* could be achieved by different approaches including 16S rRNA gene analysis and whole genome-based methods such as ANI estimation, cgMLST and core genome-phylogenetic analysis. Though these approaches were congruent, analysis of the genus core genome provided a more robust phylogenetic framework for accurate delineation of the species'

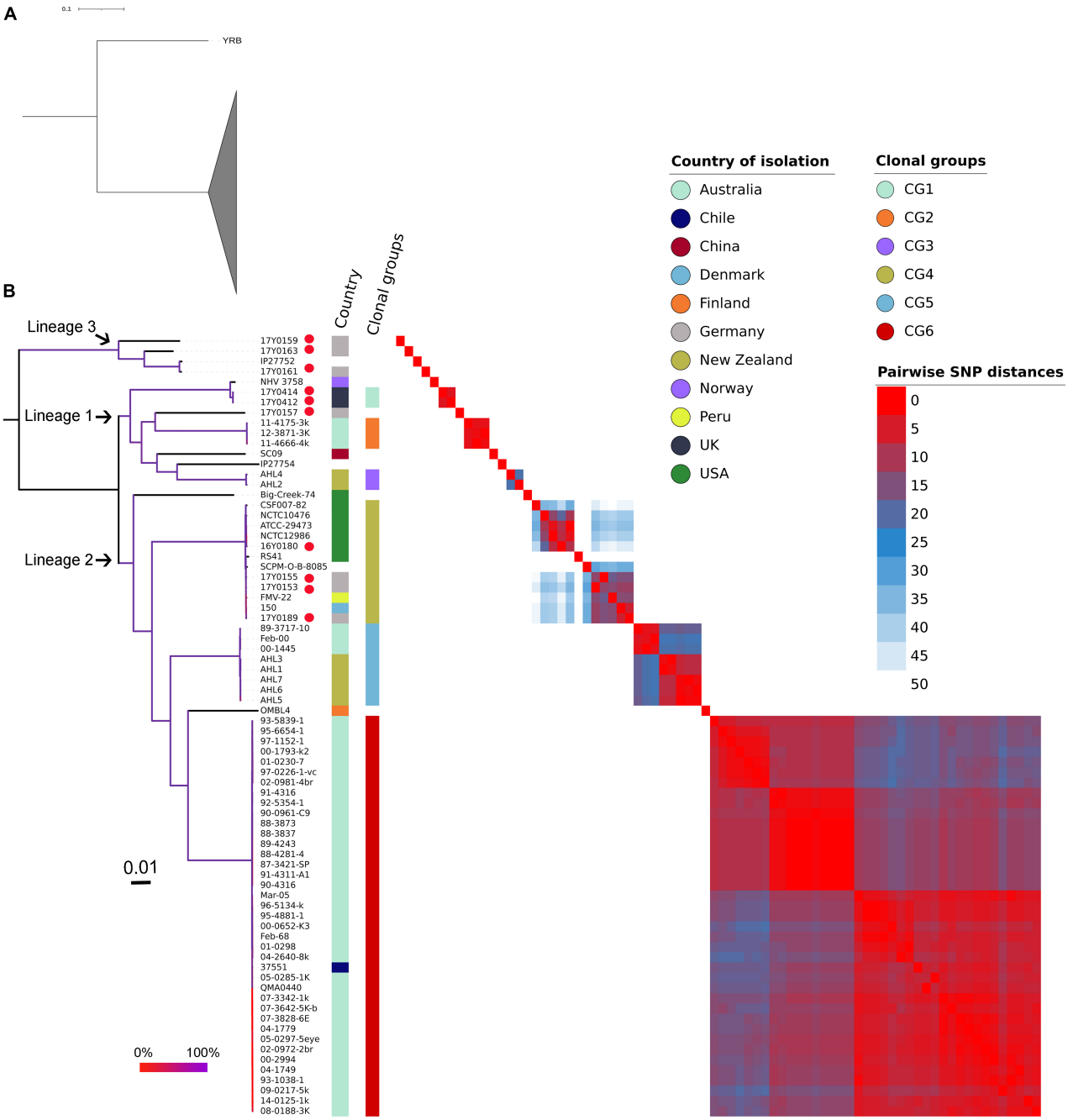


FIGURE 4 | Phylogenetic relationships of the *Yersinia ruckeri* genomes based on whole-genome SNPs after filtering regions with high SNP density using Gubbins. **(A)** Maximum likelihood phylogenetic analysis based on 77 strains showing strain YRB as the most distant to the other investigated *Y. ruckeri* genomes presented with collapsed clade. **(B)** Maximum likelihood phylogenetic tree based on 76 *Y. ruckeri* strains including the 10 strains (red dot) sequenced in this study. The three lineages described in the text are highlighted. Plotted next to the phylogenetic tree are the country of isolation, the clonal groups identified based on less than 50 pairwise SNP, and a heat map showing the SNP distances between each pair of genomes after filtering possibly recombinant regions. The branching pattern was generated by the maximum likelihood method as implemented in the RAxML program. Branch coloration refers to the bootstrap support based on the analysis of 100 resampled trees.

genetic relatedness. The large core-nucleotide genome alignment (∼2 Mb) predicted with the Panaroo pipeline provided a matrix for the super resolution phylogenetic analysis of the genus *Yersinia*. Analysis of the classical 16S rRNA marker did not correctly identify the closest relatives to *Y. ruckeri* necessitating careful interpretation of 16S rRNA results. Our results also showed that genome alignments and phylogenetic analyses are more informative and discriminatory than the

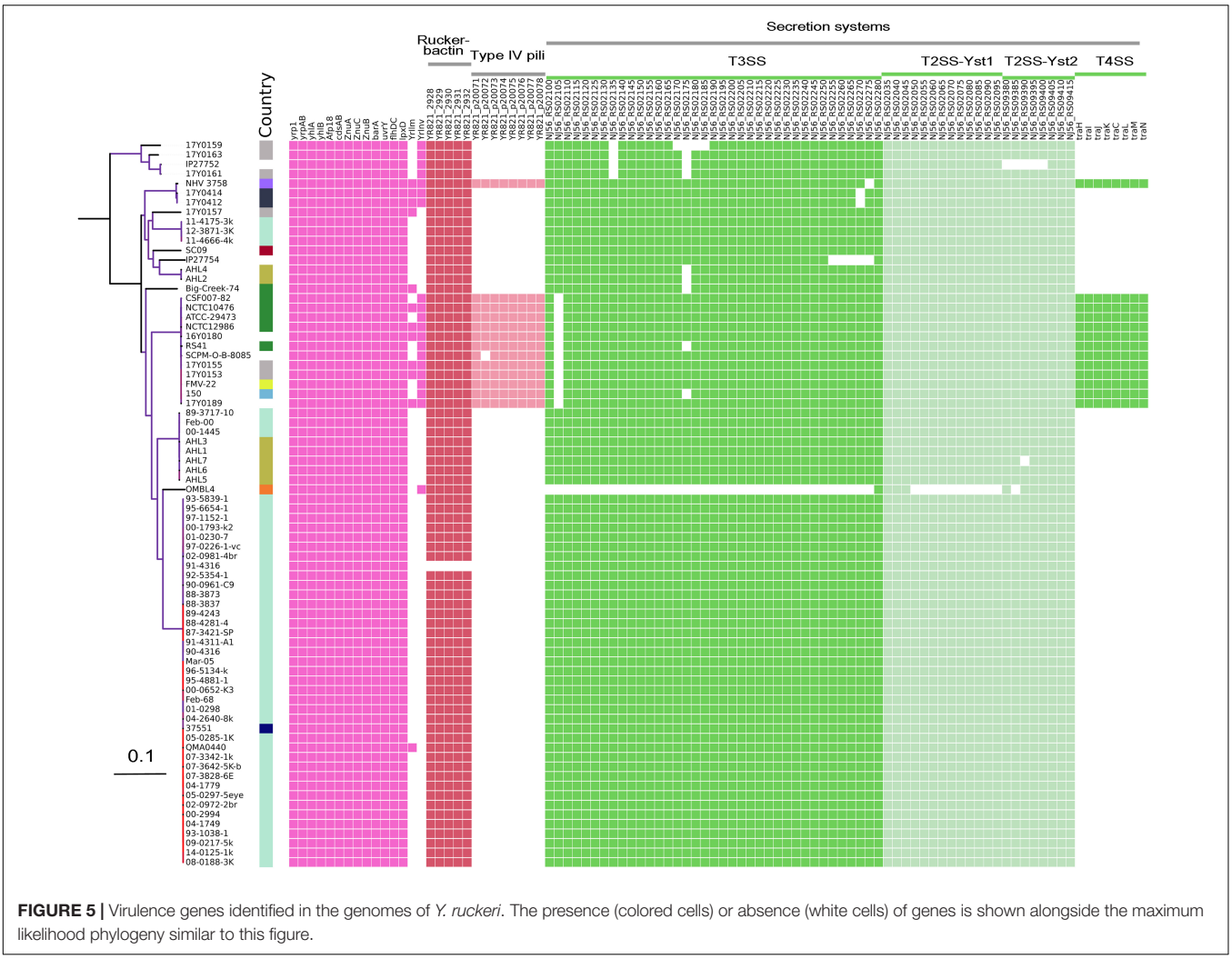


FIGURE 5 | Virulence genes identified in the genomes of *Y. ruckeri*. The presence (colored cells) or absence (white cells) of genes is shown alongside the maximum likelihood phylogeny similar to this figure.

whole genome allele typing system implemented in the genus-wide cgMLST scheme (Savin et al., 2019). Despite the fact that cgMLST provides portable isolate data easily exchangeable via a centralized database, the application of cgMLST allele typing to define species boundaries can be challenging. The *Yersinia* genus-wide scheme comprises a limited number of loci (500 genes), hence genetic diversity and true inter-strain distances are not well represented above certain thresholds, which may cause problems. The cgMLST system is very sensitive to assembly artifacts by falsely increasing allelic distances. Our results show that the phylogenetic analysis of core genome alignments is a robust approach that complements the average nucleotide identity results (Chung et al., 2018). The combination of both methods provides a stable scheme for delineating *Yersinia* species. The sole application of allele typing or 16S rRNA gene marker analysis requires caution during data interpretation.

Genomic analyses of bacterial pathogens revealed variable evolutionary processes that possibly influenced bacterial lifestyle and led to specialization in specific ecological niches. *Y. ruckeri* has arguably been discussed as a fish-restricted pathogen but the underlying genetic mechanisms and signatures of adaptation remain to be elucidated. In many bacteria, expansions of repetitive insertion sequences, large-scale chromosomal rearrangements and the high frequency of pseudogenes have been associated with narrow niche specialization as seen in e.g., *Shigella* (Feng et al., 2011), *Streptococcus agalactiae* (Almeida et al., 2016), and *Y. pestis* (Reuter et al., 2014). Although genomic analyses on the structure and layout of genes in *Y. ruckeri* revealed conservation and low incidence of chromosomal rearrangements, *Y. ruckeri* showed various mobile genetic elements. Moreover, *Y. ruckeri* displayed pseudogenization in the form of gene fragmentation and truncation most likely due to internal stop codons. These genetic changes possibly disrupt gene transcription and translation, contributing to genome streamlining via removal of genes whose functions are no longer required (Feng et al., 2011). However, in *Y. ruckeri*, pseudogenes and insertion sequences were not as pronounced as in other *Yersinia* spp. notably *Y. pestis* (Chen et al., 2010), indicative of a different evolutionary pathway. The small genome size of *Y. ruckeri* relative to other *Yersinia* species is similar to other aquatic bacteria with reduced functional repertoire. Previous studies corroborate our findings, e.g., the absence of genes

encoding for urease, methionine salvage, B12-related metabolism and myo-inositol degradation in *Y. ruckeri* (Chen et al., 2010; Tan et al., 2016).

The pattern of gene presence/absence in the accessory genome was in close agreement with the core genome phylogeny, possibly indicating concurrent co-evolution of *Yersinia* core genomes with acquisition and loss of genes. This is in agreement with a previous study (Tan et al., 2016) describing limited influence of lateral gene transfer on *Yersinia* phylogeny. In our study, we performed a GWAS analysis to identify differentially present genes in 100 representative *Yersinia* genomes. Although the selected genomes represented all validated *Yersinia* species, they may not have captured the diversity of this genus. Nevertheless, our analyses revealed the distinctive presence of 303 additional genes in *Y. ruckeri*. These genes were predicted to encode (1) virulence factors such as Afp18, Yrp1 and phage proteins and (2) metabolic enzymes involved in the utilization of iron, amino acids, specifically arginine and histidine, as well as carbohydrates. These factors are involved in the pathogenesis of ERM. Most of these genes were arranged in cassettes, which may prevent random dispersal throughout the genome. The association of these cassettes with mobile elements are indicative of horizontal acquisition. These results could indicate that a process of gene acquisition may have contributed to the evolution of the species.

Genome Comparison of *Y. ruckeri* Strains

The genome of *Y. ruckeri* is highly conserved in terms of gene structure, gene layout and functional categorization of genes. It harbors a large core genome, representing 85% of the total number of genes. Similar to other *Yersinia* species (Tan et al., 2016), diversification of the species core genome was mainly due to gene mutations. The pangenome is open with an estimated contribution of 11 new genes at the 77th sequenced genome (**Supplementary Figure 4B**). Our results corroborate and consolidate the findings of Barnes et al. (2016) by adding further strains to the analysis (total = 77) and the calculation of proposed regression models (Tettelin et al., 2008). It is very likely that the pangenome variation has occurred due to the variable mobile elements and plasmids in the strains. *Y. ruckeri* is thought to be vulnerable to the acquisition of genetic elements due to the absence of CRISPR systems (Barnes et al., 2016).

Recent studies based on whole genome sequence analysis and multilocus variable-number tandem-repeat analysis (Barnes et al., 2016; Gulla et al., 2018) described geographical endemism for numerous genetically distant lineages of *Y. ruckeri* strains. To date a large-scale analysis on *Y. ruckeri* strains from Germany is missing. However, our results allocate the German strains to three genetically distant phylogenetic lineages of the species. A possible explanation for the circulation of different lineages in Germany might be the acquisition of breeding stock and fish spawn from different geographic origins; it may also explain the presence of German strains (n = 3) in the clonal group (CG4) together with strains from the United States, Peru, and Denmark. Gulla et al. (2018) made similar observations using MLVA, pointing to a possible spread of *Y. ruckeri* strains from North America to Europe and South America.

The chromosomal background of this CG4 might have conferred a selective advantage toward acquisition and maintenance of virulence plasmids. CG4 contains a 103.8 kb large plasmid similar to the pYR3 plasmid previously described (Nelson et al., 2015). This plasmid and the 80.8 kb plasmid (pYR4) of strain NVH_3758 belong to the IncFII plasmid family (Méndez et al., 2009; Wrobel et al., 2018b); it contains several mobile protein elements and a conjugative system. A large region of ∼55 kb is present in both plasmids including a *pil* operon encoding for a type IV pilus and several *tra* operons encoding for the T4SS. The *tra* operon is important for the virulence of *Y. ruckeri* as previously demonstrated (Méndez et al., 2009). The association between the type of virulence plasmid and phylogenetic group can be explained by the clonal expansion of hypervirulent strains as a result of horizontal plasmid acquisition. On the other hand, our results show that similar virulence genes are present in the majority of the analyzed *Y. ruckeri* strains. Previous studies indicated at differences in ERM pathogenicity between *Y. ruckeri* strains (Guijarro et al., 2018). This could imply a multifactorial nature of ERM, in which environmental factors may be involved and influence the transcription and regulation of virulence genes. It can also be hypothesized that strain fitness in terms of its metabolic repertoire may have an important role in the pathogenicity of ERM in fish.

CONCLUSION

This study provides valuable insights into the genomic variability and phylogenetic relatedness of *Y. ruckeri*. Our results show that the genetic relatedness of the species is better delineated using core genome alignment followed by phylogenetic analysis rather than genome-wide allele typing or 16S rRNA gene analysis. Phylogenetically, *Y. ruckeri*, *Y. entomophaga*, and *Y. nurmii* form a species complex that is distinct from other *Yersinia* species. *Y. ruckeri* has unique genomic regions that might play a role in the pathogenesis of ERM. *Y. ruckeri* show highly homogeneous genomes with a high degree of synteny and contain numerous genomic islands but lack the CRISPR-Cas system. Limited genomic and virulence plasticity was observed arising from variable components of mobile elements and plasmids. Strain grouping based on the ML phylogeny revealed the absence of geographic clustering of several *Y. ruckeri* strains indicating a possible dissemination of a specific strain group. The German *Y. ruckeri* strains belong to the three genetically distant phylogenetic lineages.

MATERIALS AND METHODS

Cultivation and Identification of *Y. ruckeri* Strains

Strains were routinely cultured on Columbia blood agar at 25°C. Initial identification was done using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS; Bruker Daltonik GmbH, Bremen, Germany) and API 20E (Biomerieux, Nürtingen, Germany). Typing was done by Pulsed Field Gel Electrophoresis (PFGE) with *Not*-I,

Dice similarity coefficient and UPGMA. In addition, repetitive sequence-based PCR assays were performed, including BOX-AIR-based repetitive extragenic palindromic-PCR (BOX-PCR), (GTG)5-PCR, enterobacterial repetitive intergenic consensus (ERIC-PCR), and repetitive extragenic palindromic PCR (REP-PCR) (Huang et al., 2013).

Y. ruckeri Genome Sequencing

Genomic DNA was extracted using the QIAGEN Genomic-tip 100/G kit and the Genomic DNA Buffer Set (QIAGEN GmbH, Hilden, Germany) following the manufacturer's instructions. DNA quality was assessed with a Qubit 3 Fluorometer using the Qubit™ dsDNA HS Assay Kit (Invitrogen™, Germany). Whole genome sequencing was performed on the Pacific Biosciences RS sequencer with SMRT Technology PacBio RS II with SMRT Technology PacBio RS II (Pacific Biosciences, Menlo Park, CA, United States) at GATC (Konstanz, Germany), using standard protocols according to the manufacturer's instructions. Paired-end sequencing was performed using an Illumina MiSeq instrument. Sequencing libraries were created using the Nextera XT DNA Library Preparation Kit (Illumina Inc., United States) to generate reads of 300 bp in length according to the manufacturer's instructions.

Genome Assembly, Circularization and Polishing

De novo genome assembly and circularization for long sequencing reads was performed as previously described (Abdel-Glil et al., 2021). Briefly, RS_HGAP_Assembly v3 available via SMRT Analysis system v2.3.014 was used for the assembly of the PacBio data (Chin et al., 2013). Gepard v1.40 was then used to identify similar parts at the ends of each contig (Krumsiek et al., 2007). Genome circularization was performed by merging overlapping ends of each genome using Circlator v1.5.0 or check_circularity.pl from SPRAI (Hunt et al., 2015). Quiver algorithm (RS.Resequencing.1) was iteratively used for error corrections in the merged region (Chin et al., 2013). For additional polishing of the genome assemblies, Pilon v1.23 was applied four times using the Illumina data to correct the final assembled data with standard settings (Walker et al., 2014).

Genome Annotation and Functional Categorization of Genes

Assembled genomes were annotated using the rapid genome annotation pipeline Prokka v1.13.3 (Seemann, 2014). For prediction of CRISPR sequences, the CRISPR Recognition Tool v1.1 (Bland et al., 2007) was used in default mode. Prophage sequences were predicted by querying contigs assembled to the prophage databases implemented in the PHASTER web service (Arndt et al., 2016). Predicted prophages were classified as intact, questionable or incomplete using the scoring system described by Arndt et al. (2016). For prediction of insertion elements, ISEscan v1.7.2 software (Xie and Tang, 2017) was used with the "–removeShortIS" flag enabled to report only complete IS elements, i.e., IS elements with terminal inverted repeat sequences and longer than 400 bp. The circularized *Y. ruckeri* plasmids were separated and analyzed for their degree of homology by applying pyani v0.2.3 software (Pritchard et al., 2016) to report the average nucleotide identity values, percentage of genome aligned and the Hadamard distance matrix. The latter merges the average nucleotide identity values and the percentage of the aligned genomic regions. Pyani implements alignment algorithms for ANI calculation. Here, the BLAST algorithm was used for alignment.

Reference based identification of potential pseudogenes was done with Pseudofinder v1.0[1]. The non-redundant NCBI protein database was used as reference database for Pseudofinder, and diamond 0.9.24 (Buchfink et al., 2015) for high throughput protein alignment. Functional annotation of genes using orthology assignment was performed using eggNOG-mapper v2.0 (Huerta-Cepas et al., 2018). Chromosome alignment of orthologous and xenologous regions of the *Y. ruckeri* strains was performed using progressive Mauve algorithm applied in standard mode (Darling et al., 2010).

Genome Comparison and Phylogenetic Analysis of the *Yersinia* Genus

For genus-wide comparison, barrnap v0.9[2] was used to predict and extract the rRNA genes. Extracted 16S rRNA sequences were aligned using MAFFT v7.307 (Katoh and Standley, 2013) followed by an ML phylogenetic analysis using RAxML v8.2.10 (Stamatakis, 2014) with the general time-reversible (GTR)-gamma model and 100 bootstrap replicates.

Estimation of the gANI between the 100 *Yersinia* genomes was done using pyani v0.2.3 (Pritchard et al., 2016), using standard parameters. ANI values were estimated by applying two independent methods: FastANI v1.3 which uses the kmer content for alignment-free sequence mapping (Jain et al., 2018), and the MiSI (Varghese et al., 2015) tool "ANIcalculator v1," which excludes RNA genes and reports alignment fractions and ANI values based on the orthologous protein-coding genes identified as bidirectional best hits.

To identify the genus core genome, we used Panaroo v1.2.7 to identify the core orthologous genes of the 100 *Yersinia* genomes using the stringency mode "strict" (Tonkin-Hill et al., 2020). The default thresholds of Panaroo were applied including sequence identity at 98% and core genome assignment at 95%. Gene-by-gene alignment was done with MAFFT v7.307 as in Panaroo. Concatenated core gene alignment was then used for ML phylogenetic analysis using RAxML v8.2.10 as mentioned above.

The cgMLST allele typing for the 100 *Yersinia* strains was predicted by submitting the genome sequences to the *Yersinia* cgMLST scheme using the Pasteur MLST database[3] (Savin et al., 2019). Next, the Genome Comparator plugin of BIGSdb implemented in the Pasteur database was used for genome comparison. The resulting allelic profiles were visualized in a minimum spanning tree using the online version of Grapetree software (Zhou et al., 2018).

[1]https://github.com/filip-husnik/pseudofinder

[2]https://github.com/tseemann/barrnap

[3]https://bigsdb.pasteur.fr/yersinia/

Pan-GWAS analysis was done with Scoary v1.6.10 (Brynildsrud et al., 2016) to identify the statistical association between the genus accessory genes and the *Y. ruckeri* species. We used the accessory genes defined with the Panaroo program as input for Scoary. A total of 10,000 permutation replicates was applied using genes with 100% sensitivity and specificity, and with *P*-values of less than 10^{-5} (naïve, Benjamini–Hochberg-adjusted and Bonferroni). To confirm the specificity of the identified genes for *Y. ruckeri*, we applied BLAST analysis to the nucleotide sequences at 90% sequence identity and 80% coverage. Functional annotation of the genes was done using the COG database as described above. Furthermore, PSORTb version 3.0.3 was used for subcellular localization prediction (Yu et al., 2010).

Genome Comparison and Phylogenetic Analysis for *Y. ruckeri* Strains

For species-wide comparison, 77 *Y. ruckeri* genomes were included. Pangenome analysis was performed as mentioned above using the Panaroo pipeline (Tonkin-Hill et al., 2020). PanGP v1.0.1 program (Zhao et al., 2014) was used for the analysis of pangenome profiles with a distance guide algorithm of 100 replicates and 1000 permutations of genome order.

Parsnp v1.2 (Treangen et al., 2014) performed a core genome alignment with default settings. Filtering regions with high SNP density from the alignment was done with Gubbins v2.4.1 in standard mode (Croucher et al., 2015). RAxML v8.2.10 was applied as mentioned above to construct ML phylogenetic trees (Stamatakis, 2014). Phylogenetic trees were visualized using iTOLv6 (Letunic and Bork, 2019).

A search for the virulence related genes in *Y. ruckeri* was performed using BLASTN v2.10.1+ (Altschul et al., 1997) as featured in ABRicate v1.0.1[4]. A sequence identity threshold of 80% and coverage threshold of 80% was applied.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. These data can be found in the NCBI under BioProject PRJNA767598 (https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA767598).

AUTHOR CONTRIBUTIONS

MA-G: bioinformatic data analyses and manuscript draft. UF, UM, DS, and LS: cultivation and microbial analysis of *Y. ruckeri* and manuscript revision. HN and LS: project planning, data analysis, and manuscript revision. All authors contributed to the article and approved the submitted version.

[4] https://github.com/tseemann/abricate

FUNDING

This work was funded in house by the Friedrich-Loeffler-Institut.

ACKNOWLEDGMENTS

The excellent technical assistance of Julia Haller, Claudia Grosser, and Johannes Solle at the Friedrich-Loeffler-Institut, IBIZ is greatly appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.782415/full#supplementary-material

Supplementary Figure 1 | Representation of the circular chromosome of the *Yersinia ruckeri* type strain (3,696,254 bp). Circles inward represent predicted annotated coding genes in clockwise (1) and anticlockwise (2) orientations with genes colored according to the predicted clusters of orthologous classes as shown in the legend. Circle 3 in black represents the GC content. Circle 4 shows histogram plots for positive (purple) and negative (green) GC skews. Positive skew indicates a higher presence of G than C while negative skew indicates more C than G based on the equation (G−C)/(G+C) calculated over a window of 10 kb.

Supplementary Figure 2 | Phylogenetic relationship within the *Yersinia* genus based on 16S rRNA sequences. The branching pattern was generated by the maximum likelihood method as implemented in the RAxML program. The tree is midpoint rooted. Branch coloration refers to the bootstrap support from zero (red) to 100 (violet) based on the analysis of 100 resampled trees. Collapsed nodes include taxa of a single species.

Supplementary Figure 3 | Results of the average nucleotide identity (ANI) analysis. **(A)** A heat map showing the ANI values calculated over the base pair matches of the aligned regions between each pair of genomes using pyani (Pritchard et al., 2016); **(B)** Histogram of the ANI values between each genome pair produced with bactaxR (Carroll et al., 2020); **(C)** Heat map showing the alignment coverage for the calculation of nucleotide identity; the dendrograms shown above and on the left side of the heat maps are based on hierarchical clustering of ANI values using the simple linkage method as described (Pritchard et al., 2016).

Supplementary Figure 4 | Pangenome analysis encompassing 77 *Yersinia ruckeri* genomes. **(A)** Plot showing the trajectory pattern of expansion of pangenome genes (blue curve) versus the reduction of the core genome (green curve). The plot depicts the total number of pan and core genes per each sequenced genome. **(B)** Plot representing the contribution of new genes to the overall gene pool with sequential addition of genomes.

REFERENCES

Abdel-Glil, M. Y., Thomas, P., Linde, J., Busch, A., Wieler, L. H., Neubauer, H., et al. (2021). Comparative in Silico genome analysis of *Clostridium perfringens* unravels stable phylogroups with different genome characteristics and pathogenic potential. *Sci. Rep.* 11:6756. doi: 10.1038/s41598-021-86148-8

Almeida, A., Alves-Barroco, C., Sauvage, E., Bexiga, R., Albuquerque, P., Tavares, F., et al. (2016). Persistence of a dominant bovine lineage of group B *Streptococcus* reveals genomic signatures of host adaptation. *Environ. Microbiol.* 18, 4216–4229. doi: 10.1111/1462-2920.13550

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

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Barnes, A. C., Delamare-Deboutteville, J., Gudkovs, N., Brosnahan, C., Morrison, R., and Carson, J. (2016). Whole genome analysis of *Yersinia ruckeri* isolated over 27 years in Australia and New Zealand reveals geographical endemism over multiple lineages and recent evolution under host selection. *Microb. Genomics* 2:e000095. doi: 10.1099/mgen.0.000095

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17:238.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Carroll, L. M., Wiedmann, M., and Kovac, J. (2020). Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. *mBio* 11:e00034-e20. doi: 10.1128/mBio.00034-20

Chen, P. E., Cook, C., Stewart, A. C., Nagarajan, N., Sommer, D. D., Pop, M., et al. (2010). Genomic characterization of the *Yersinia* genus. *Genome Biol.* 11:R1.

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Chung, M., Munro, J. B., Tettelin, H., and Dunning Hotopp, J. C. (2018). Using core genome alignments to assign bacterial species. *mSystems* 3:e00236-18.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/nar/gku1196

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147

Feng, Y., Chen, Z., and Liu, S.-L. (2011). Gene decay in *Shigella* as an incipient stage of host-adaptation. *PLoS One* 6:e27754. doi: 10.1371/journal.pone.0027754

Guijarro, J. A., García-Torrico, A. I., Cascales, D., and Méndez, J. (2018). The infection process of *Yersinia ruckeri*: reviewing the pieces of the Jigsaw puzzle. *Front. Cell. Infect. Microbiol.* 8:218. doi: 10.3389/fcimb.2018.00218

Gulla, S., Barnes, A. C., Welch, T. J., Romalde, J. L., Ryder, D., Ormsby, M. J., et al. (2018). Multi-Locus Variable number of tandem repeat Analysis (MLVA) of *Yersinia ruckeri* confirms the existence of host-specificity, geographic endemism and anthropogenic dissemination of virulent clones. *Appl. Environ. Microbiol.* 84:e00730-18. doi: 10.1128/AEM.00730-18

Huang, Y., Runge, M., Michael, G. B., Schwarz, S., Jung, A., and Steinhagen, D. (2013). Biochemical and molecular heterogeneity among isolates of *Yersinia ruckeri* from rainbow trout (*Oncorhynchus mykiss*, Walbaum) in North West Germany. *BMC Vet. Res.* 9:215. doi: 10.1186/1746-6148-9-215

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2018). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 16:294. doi: 10.1186/s13059-015-0849-0

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026–1028. doi: 10.1093/bioinformatics/btm039

Kumar, G., Menanteau-Ledouble, S., Saleh, M., and El-Matbouli, M. (2015). *Yersinia ruckeri*, the causative agent of enteric redmouth disease in fish. *Vet. Res.* 46:103. doi: 10.1186/s13567-015-0238-4

Le Guern, A.-S., Savin, C., Angermeier, H., Brémont, S., Clermont, D., Mühle, E., et al. (2020). *Yersinia artesiana* sp. nov., *Yersinia proxima* sp. nov., *Yersinia alsatica* sp. nov., *Yersina vastinensis* sp. nov., *Yersinia thracica* sp. nov. and *Yersinia occitanica* sp. nov., isolated from humans and animals. *Int. J. Syst. Evol. Microbiol.* 70, 5363–5372. doi: 10.1099/ijsem.0.004417

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Méndez, J., Fernández, L., Menéndez, A., Reimundo, P., Pérez-Pascual, D., Navais, R., et al. (2009). A chromosomally located traHIJKCLMN operon encoding a putative type IV secretion system is involved in the virulence of *Yersinia ruckeri*. *Appl. Environ. Microbiol.* 75, 937–945. doi: 10.1128/AEM.01377-08

Nelson, M. C., Lapatra, S. E., Welch, T. J., and Graf, J. (2015). Complete genome sequence of *Yersinia ruckeri* strain CSF007-82, etiologic agent of red mouth disease in Salmonid fish. *Genome Announc.* 3:e01491-14. doi: 10.1128/genomeA.01491-14

Nguyen, S. V., Cunningham, S. A., Jeraldo, P., Tran, A., and Patel, R. (2021). *Yersinia occitanica* is a later heterotypic synonym of *Yersinia kristensenii* subsp. rochesterensis and elevation of *Yersinia kristensenii* subsp. rochesterensis to species status. *Int. J. Syst. Evol. Microbiol.* 71:4626. doi: 10.1099/ijsem.0.004626

Ormsby, M. J., and Davies, R. L. (2021). Diversification of OmpA and OmpF of *Yersinia ruckeri* is independent of the underlying species phylogeny and evidence of virulence-related selection. *Sci. Rep.* 11:3493. doi: 10.1038/s41598-021-82925-7

Ormsby, M. J., Grahame, E., Burchmore, R., and Davies, R. L. (2019). Comparative bioinformatic and proteomic approaches to evaluate the outer membrane proteome of the fish pathogen *Yersinia ruckeri*. *J. Proteomics* 199, 135–147. doi: 10.1016/j.jprot.2019.02.014

Paradiso, R., Orsini, M., Bolletti Censi, S., Borriello, G., and Galiero, G. (2016). Complete genome sequence of a *Myoviridae* bacteriophage infecting *Salmonella enterica* Serovar Typhimurium. *Genome Announc.* 4:e00940-16. doi: 10.1128/genomeA.00940-16

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/c5ay02550h

Reuter, S., Connor, T. R., Barquist, L., Walker, D., Feltwell, T., Harris, S. R., et al. (2014). Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6768–6773. doi: 10.1073/pnas.1317161111

Richter, M., and Rossello-Mora, 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

Romalde, J. L., and Toranzo, A. E. (1993). Pathological activities of *Yersinia ruckeri*, the enteric redmouth (ERM) bacterium. *FEMS Microbiol. Lett.* 112, 291–299. doi: 10.1111/j.1574-6968.1993.tb06465.x

Savin, C., Criscuolo, A., Guglielmini, J., Le Guern, A.-S., Carniel, E., Pizarro-Cerdá, J., et al. (2019). Genus-wide Yersinia core-genome multilocus sequence typing for species identification and strain characterization. *Microb. Genomics* 5:e000301. doi: 10.1099/mgen.0.000301

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tan, S. Y., Tan, I. K. P., Tan, M. F., Dutta, A., and Choo, S. W. (2016). Evolutionary study of Yersinia genomes deciphers emergence of human pathogenic species. *Sci. Rep.* 6:36116. doi: 10.1038/srep36116

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006

Tonkin-Hill, G., Macalasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21:180. doi: 10.1186/s13059-020-02090-4

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/s13059-014-0524-x

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., et al. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 43, 6761–6771.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wrobel, A., Leo, J. C., and Linke, D. (2019). Overcoming fish defences: the virulence factors of *Yersinia ruckeri*. *Genes (Basel)* 10:700. doi: 10.3390/genes10090700

Wrobel, A., Ottoni, C., Leo, J. C., Gulla, S., and Linke, D. (2018a). The repeat structure of two paralogous genes, *Yersinia ruckeri* invasin (yrInv) and a "Y. ruckeri invasin-like molecule", (yrIlm) sheds light on the evolution of adhesive capacities of a fish pathogen. *J. Struct. Biol.* 201, 171–183. doi: 10.1016/j.jsb.2017.08.008

Wrobel, A., Ottoni, C., Leo, J. C., and Linke, D. (2018b). pYR4 from a Norwegian isolate of *Yersinia ruckeri* is a putative virulence plasmid encoding both a type IV Pilus and a type IV secretion system. *Front. Cell. Infect. Microbiol.* 8:373. doi: 10.3389/fcimb.2018.00373

Wrobel, A., Saragliadis, A., Pérez-Ortega, J., Sittman, C., Göttig, S., Liskiewicz, K., et al. (2020). The inverse autotransporters of *Yersinia ruckeri*, YrInv and YrIlm, contribute to biofilm formation and virulence. *Environ. Microbiol.* 22, 2939–2955. doi: 10.1111/1462-2920.15051

Xie, Z., and Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33, 3340–3347. doi: 10.1093/bioinformatics/btx433

Yang, H., Zhujin, D., Marana, M. H., Dalsgaard, I., Rzgar, J., Heidi, M., et al. (2021). Immersion vaccines against *Yersinia ruckeri* infection in rainbow trout: comparative effects of strain differences. *J. Fish Dis.* 1–14. doi: 10.1111/jfd.13507

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., et al. (2014). PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics (Oxf. Engl.)* 30, 1297–1299. doi: 10.1093/bioinformatics/btu017

Zhou, Z., Alikhan, N.-F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 January 2022
doi: 10.3389/fmicb.2022.826894

Experimental Evolution Expands the Breadth of Adaptation to an Environmental Gradient Correlated With Genome Reduction

*Masaomi Kurokawa, Issei Nishimura and Bei-Wen Ying**

School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Whether and how adaptive evolution adjusts the breadth of adaptation in coordination with the genome are essential issues for connecting evolution with ecology. To address these questions, experimental evolution in five *Escherichia coli* strains carrying either the wild-type genome or a reduced genome was performed in a defined minimal medium (C0). The ancestral and evolved populations were subsequently subjected to fitness and chemical niche analyses across an environmental gradient with 29 combinations of eight chemical components of the minimal medium. The results showed that adaptation was achieved not only specific to the evolutionary condition (C0), but also generally, to the environmental gradient; that is, the breadth of adaptation to the eight chemical niches was expanded. The magnitudes of the adaptive improvement and the breadth increase were both correlated with genome reduction and were highly significant in two out of eight niches (i.e., glucose and sulfate). The direct adaptation-induced correlated adaptation to the environmental gradient was determined by only a few genome mutations. An additive increase in fitness associated with the stepwise fixation of mutations was consistently observed in the reduced genomes. In summary, this preliminary survey demonstrated that evolution finely tuned the breadth of adaptation correlated with genome reduction.

Keywords: experimental evolution, niche, local adaptation, global adaptation, growth fitness, culture medium, experimental ecology

OPEN ACCESS

Edited by:
Daniel Yero, Universidad Autónoma de Barcelona, Spain

Reviewed by:
Angus Buckling, University of Exeter, United Kingdom
Nishad Matange, Indian Institute of Science Education and Research, Pune, India

***Correspondence:**
Bei-Wen Ying
ying.beiwen.gf@u.tsukuba.ac.jp

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

Received: *01 December 2021*
Accepted: *06 January 2022*
Published: *26 January 2022*

Citation:
Kurokawa M, Nishimura I and Ying B-W (2022) Experimental Evolution Expands the Breadth of Adaptation to an Environmental Gradient Correlated With Genome Reduction. Front. Microbiol. 13:826894. doi: 10.3389/fmicb.2022.826894

INTRODUCTION

Microorganisms living in nature show highly diverse habitats (i.e., ecological niches) as a consequence of local adaptation (Kawecki and Ebert, 2004) and are constrained by evolutionary costs (Bono et al., 2020). The ecological niche is believed to be associated with genomic information (Alneberg et al., 2020), which is considered to be a result of adaptive evolution (Batut et al., 2014). Numerous studies have reported adaptation to a certain niche related to genetic causes, such as linkages between genome streamlining and niche partitioning (Graham and Tully, 2021), gene loss and niche shifts (Chu et al., 2021), genome reduction and habitat transition (Salcher et al., 2019), metabolic costs (Ankrah et al., 2018), genome architecture and habitat (Getz et al., 2018) or niche-directed evolution (Andrei et al., 2019). These findings provide strong evidence linking adaptive evolution to ecological niches in terms of the spatial and environmental differentiation of species. As environmental changes are more often gradual under temporal and spatial restrictions, whether the breadth of adaptation to environmental gradients is shaped by evolution is an intriguing question.

Adaptation to a certain environment (i.e., niche) is often investigated by means of experimental evolution (Kawecki et al., 2012; Barrick and Lenski, 2013) to acquire direct evidence and preform a precise evaluation. In general, these studies have focused on a target component among the numerous components that comprise the environment, such as carbon sources (Satterwhite and Cooper, 2015) or antibiotics (Baym et al., 2016), as the factor triggering adaptive evolution. The environment, whether it is the culture medium used in the laboratory or the ecological niche in nature, is comprised of not only the target component but also a number of other nutrients and trace elements. Thus, adaptation must occur not only to the target component but also to all of the remaining components in the environment. However, the participation of components other than the target component in adaptive evolution has generally been neglected. A machine learning analysis of medium components showed that it was the trace elements (e.g., metal ions) rather than the major nutrients (e.g., glucose) that determined bacterial growth, which was sensitive to the concentration gradient (Ashino et al., 2019). Thus, whether and how adaptation through experimental evolution is associated with correlated adaptation to environmental gradients must be addressed.

In addition, genome size, as a quantitative index of genetic richness, has been intensively studied; nevertheless, its impact on adaptation remains unclear. Changes in genome size are commonly observed in nature (Kuo and Ochman, 2009; Batut et al., 2014; Maistrenko et al., 2020) and are known as one of the driving forces of adaptive evolution (e.g., horizontal gene transfer) (Keeling and Palmer, 2008; Daubin and Szöllősi, 2016). Genome size can be experimentally reduced (Posfai et al., 2006; Kato and Hashimoto, 2007; Mizoguchi et al., 2008) to determine the minimal genetic requirement of living organisms (Xavier et al., 2014; Rees-Garbutt et al., 2020). Such reduced genomes tend to show decreased fitness (Karcagi et al., 2016; Kurokawa et al., 2016) and increased mutagenesis (Nishimura et al., 2017), which can both be restored by experimental evolution (Nishimura et al., 2017). These studies have revealed that genome reduction not only has evolutionary consequences but also plays a role in adaptation. To date, the experimental evidence of the contribution of genome reduction to adaptation is largely insufficient.

To investigate whether experimental evolution in a defined steady condition caused a change in the breadth of adaptation (**Figure 1A**), a pilot survey of the adaptation connecting genome reduction with environmental gradients was performed in the present study. Direct adaptation was achieved by experimental evolution, which was conducted with an assortment of laboratory *Escherichia coli* (*E. coli*) strains derived from the same parental wild-type genome with different genome sizes (**Figure 1B**). Whether and how direct adaptation contributes to correlated adaptation to an environmental gradient were quantitatively evaluated by fitness assays and chemical niche analysis. The impact of genome reduction on adaptation was analyzed in parallel, which provides the experimental demonstration and insight that connecting the two "unrelated" issues of correlated adaptation and genome reduction. The results filled the blank of knowledge on the breadth of adaptation and genome.

RESULTS

Adaptation Correlated With Genome Reduction

The adaptation of the genomes of different sizes was consistently achieved through experimental evolution under stable conditions. Five laboratory *E. coli* strains carrying either a wild-type (N0) or reduced (N7, 14, 20, or 28) genome (**Supplementary Table 1**) were subjected to experimental evolution in a chemically defined medium (C0). The evolution experiment was performed by serial transfer at a series of different dilution rates to maintain bacterial growth within the exponential phase (**Figure 2A** and **Supplementary Figure 1**). The daily records showed that a gradual increase in the growth rate consistently occurred in the reduced genomes (**Figure 2B**), and the population that proceeded (blue) along the dilution series (black) presented a somewhat rapid fitness increase. The evolutionary trajectories of the reduced genomes were somehow similar, in comparison to that of the wild-type genome. It was unclear whether the evolutionary path was attributed to the deletion of genomic sequences or the relatively low fitness of the ancestor population. The improved fitness of the endpoint population indicated that adaptation to C0 was achieved, which was highly significant in the lineages of reduced genomes. It should be noted that the fluctuation in the growth rate recorded in N20 and N28 was largely due to a pause of serial transfer, which was restarted from the glycerol stock of the bacterial population stored the day before.

Fitness and activity assays showed that adaptation was correlated with genome reduction. The comparison of the ancestors (Ancs) and the evolved populations (Evos) after approximately 1,000 generations (**Figure 2B**, pink) showed that the growth rates of Evos were all significantly higher than those of Ancs (**Figure 2C**). Positive correlations were observed between the growth rates and redox activities (**Figure 2D**) as well as between the changes thereof (**Figure 2E**), which verified that direct adaptation was achieved at both the growth and metabolic levels. Genome reduction-correlated changes in the growth rate were identified (**Figure 2F**). The evolutionary rate of the changes in the growth rate was correlated with genome reduction (**Figure 2G**), which was consistent with the correlation between genome reduction and the spontaneous mutation rate (Nishimura et al., 2017), a global parameter representing evolvability.

Direct Adaptation-Mediated Correlated Adaptation to the Environmental Gradient

Whether the adaptation to C0 caused adaptation or maladaptation across the environmental gradient was further evaluated. A total of 29 medium combinations (C0, C1∼28) were prepared with seven pure chemical substances that were

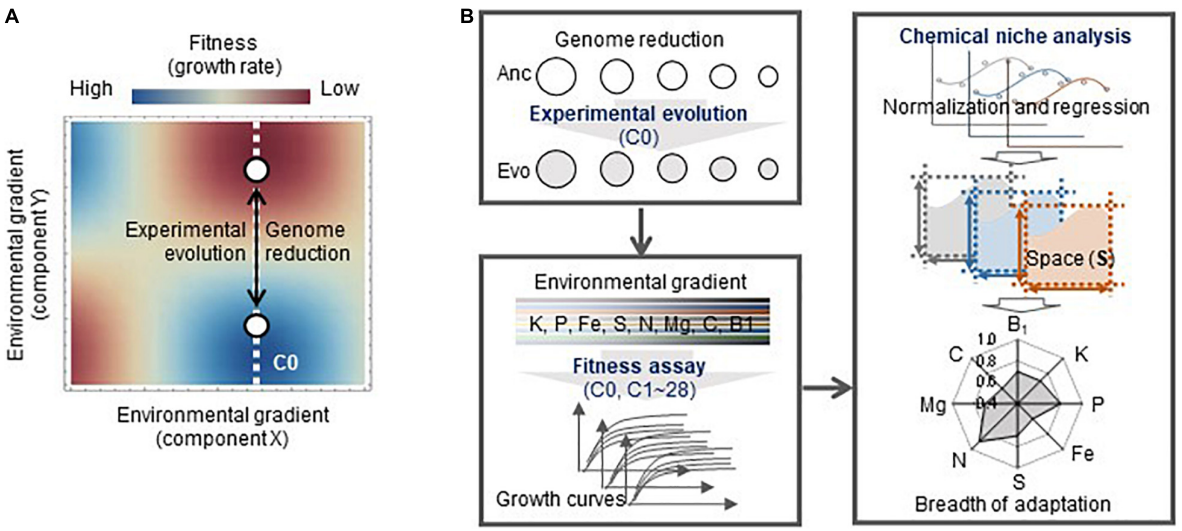


FIGURE 1 | Conceptual illustration of the study. **(A)** Fitness landscapes of the reduced genomes in the environmental gradient. The evolution of the genomes (open circles), from Anc to Evo, leads to not only adaptation to the environment for evolution (C0, white broken line) but also correlated adaptation to the environmental gradient. The color gradation from red to blue indicates the fitness from high to low. **(B)** Overview of the experiments and analyses performed in the present study. The experimental and analytical studies are categorized in three boxes, which are the experimental evolution, high-throughput fitness assay, and chemical niche analysis. Five genomes (strains) used for the evolution are shown as circles. The eight elements investigated in the chemical niche analysis are indicated.

included in C0 (**Supplementary Table 2**). These combinations comprised eight constituents (e.g., ions), whose concentrations varied broadly on a logarithmic scale (**Figure 3A**). Adaptiveness, represented by the growth rates in the exponential phase, was evaluated according to a total of 2,220 growth curves (**Supplementary Table 3**). Overall, a global increase in the growth rate across the environmental gradient of the 29 combinations was detected in the Evos with the reduced genomes (**Figure 3B**, pink) in comparison to the common decrease in the growth rate of Ancs (**Figure 3B**, green). The increased growth rates of the Evos with the reduced genomes were highly significant in most tested combinations (**Figure 3C** and **Supplementary Table 4**). This result indicated that adaptation was achieved not only specific to C0, but also globally, to the environmental gradient.

Whether and how such correlated adaptation was determined by any of the eight constituents was analyzed. The growth rates of Ancs and Evos in each constituent were plotted in response to the concentration gradient (**Figure 4**). A larger reduction in the genome was likely to be associated with larger changes in the growth rate between Anc and Evo, regardless of the variation in the constituents. In contrast, the patterns of the changes in the growth rate were dependent on the constituents. Different patterns (dynamics) among the eight constituents were consistently observed in all genomes, suggesting that the breadth of adaptation was chemically dependent, (i.e., niche dependent).

Niche-Specific and Genome Reduction-Correlated Expansion of the Breadth of Adaptation

To evaluate the breadth of adaptation, representing the equitability of fitness along the environmental gradient (Lynch and Gabriel, 1987), the space (***S***) was newly defined as the shadowed space under the fitting curve of cubic polynomial regression to the normalized dynamic of change in growth rate according to chemical concentration, in which the maxima of both the concentration gradient and the growth rate were rescaled to one unit (**Supplementary Figure 2**). Normalization and regression suggested a global parameter of ***S*** available for quantitative comparison among the varied chemical niches and genomes, where a larger ***S*** indicated a wider breadth of adaptation. A total of 80 ***S*** values (**Supplementary Figure 3**) and 40 changes in ***S*** (**Supplementary Figure 4**) were calculated accordingly. A correlation of genome reduction with the ***S*** of Ancs in the chemical niches of glucose, SO_4^{2-} and NH_4^+ was observed (**Figure 5A**, green). Additionally, a correlation of genome reduction with the evolutionary changes in ***S*** in the chemical niches of glucose and SO_4^{2-} (**Figure 5B**) was identified. The correlation with genome reduction could also account for direct adaptation (**Supplementary Figure 5**), as it was accompanied by genome reduction (**Figure 2**). The chemical niches related to carbon and sulfate were likely to be highly essential and sensitive for adaptation.

The overall improvement of the breadth of adaptation to the eight chemical niches was evaluated by the total ***S***, which was determined as the sum of the eight ***S*** values. It was narrowed in response to genome reduction but significantly enlarged due to experimental evolution (**Figure 5C**). The total ***S*** of Evos entirely encompassed that of Ancs among the bacteria with reduced genomes with larger deletions (N14, N20, and N28) but partially overlapped in the bacteria with a wild-type genome (N0) or a reduced genome with a relatively small deletion (N7). Taking N0 as an example, the increase in the ***S*** of thiamine, K^+, PO_4^{3-}, and Fe^{2+} and the decrease in the ***S*** of glucose,

Mg^{2+}, NH_4^+, and SO_4^{2-} implied that the improvement in the breadth of adaptation was chemical niche dependent and directional. In contrast, the omnidirectional expansion of total ***S*** occurred in the reduced genomes of N14, N20 and N28, which indicated general adaptation to the eight chemical niches. The adaptation to C0 restored the total ***S*** of all Evos to a level roughly equivalent to that of the wild-type genome (**Figure 5D**), indicating homeostasis in the adaptation to environmental gradients. The variation in ***S*** among the eight niches mostly declined in Evos (**Figure 5E**), indicating a balance in the breadth of adaptation to varied chemical niches. In addition, the changes in total ***S*** were positively correlated with direct adaptation and genome reduction (**Figure 5F**). These results indicated that the direct adaptation of *E. coli* expanded the breadth of adaptation to the chemical niches constituting the environment for homeostatic and balanced adaptation to the habit.

Improved Adaptiveness Attributed to Stepwise Mutation Accumulation

Genome mutation analysis (**Supplementary Table 5**) detected an approximately equivalent number of gene mutations in Evos (**Supplementary Table 6**). It should be noted that the deletion of transposons was ignored, and the mutations fixed in Evos were identified in the reduced genomes but not in the wild-type genome. The temporal changes in the allele frequency of mutants consistently showed that the mutations accumulated serially and were fixed in a stepwise manner (**Figure 6A**). Intriguingly, only a few mutations compensated for the genome reduction, independent of the degree of genome reduction. Abundant genetic information could be substituted with the modification of certain gene functions for equivalent adaptiveness, providing intriguing insight into genetic requirements. The findings were consistent with a previous report that a few mutations could cause the metabolic rewiring of a reduced genome (Choe et al., 2019). The majority of mutated genes were related to transporters and regulators, which indicated that resource diffusion for utilization and global gene regulation contributed to adaptation to the environmental gradient.

How the stepwise accumulation of the mutations contributed to adaptation was further investigated. The mutants carrying mutations in the order of their evolutionary accumulation were successfully acquired by single-colony isolation. A gradual increase in the growth rate of the mutants in the order of mutation accumulation was generally observed (**Figure 6B**), except for a transient decrease caused by the second mutation that occurred in N14. This demonstrated that the mutations were beneficial and contributed to adaptation in an additive manner. Notably, mutants with the second mutation (*rbsR* and *fliE/fliF*) in N7 failed to be acquired, indicating the co-fixation of the second mutation (*fliE/fliF*) and third mutation (*trkH*) during evolution. Intriguingly, a negative correlation between the growth rate and degree of fitness increase due to mutation was observed; that is, lower growth rates prior to mutation fixation led to greater changes in the growth rate after the mutation was fixed (**Figure 6C**). The first mutations were more likely to improve growth fitness than the mutations that were fixed later, although the statistical significance was weak (**Supplementary Figure 6**) because there were too few mutations. This finding agreed well with the rule of declining adaptability (Couce and Tenaillon, 2015) and the predictivity of the mutation-mediated fitness landscape (de Visser and Krug, 2014; Fragata et al., 2019).

Proposed Mechanism of Direct Adaptation-Induced Niche Expansion

Fitness landscape analysis (Tenaillon, 2014; Martin and Lenormand, 2015), which is applied to explain mutation occurrence (Nahum et al., 2015; Bajic et al., 2018) with a change in distance to the fitness peak (Barrick et al., 2010; Schick et al., 2015) as an evolutionary constraint (Szamecz et al., 2014; Nahum et al., 2015), was employed to explain the present findings (**Figure 7A**). According to previous reports (Konstantinidis and Tiedje, 2004; Kurokawa et al., 2016), larger deletions in the genome lead to a greater distance from the fitness peak (**i**). Since direct adaptation was achieved in correlation with genome reduction (**Figure 2**), a greater distance from the fitness peak meant that a larger change was required to achieve equivalent fitness (**ii**). The fitness increase was additive owing to the stepwise fixation of gene mutations (**Figure 6**). Since direct adaptation expanded the breadth of adaptation to the environmental gradient (**Figures 3–5**), the location in the initial fitness landscape (e.g., distance from peak of C_0) likely determined the adaptiveness to the alternative environmental gradient (e.g., C_N) (**iii**). That is, there was a higher probability of an adaptive trade-off in C_N when Anc was located closer to the adaptive peak of C_0 and a higher probability of correlated adaptation when Anc was located farther from the adaptive peak of C_0. This mechanism was consistent with the pleiotropic costs of carbon utilization found in the experimental population (Jasmin and Zeyl, 2013).

Locational bias in the initial fitness landscape (**Figure 7A**) might cause so-called preadaptation (Cullum et al., 2001) to alternative environments. A weak but statistically significant correlation of ***S*** was detected between Ancs and Evos across the genomes and chemical niches (**Figure 7B**). The changes in ***S*** were highly significantly correlated with the ***S*** of Ancs but not with those of Evos (**Figure 7C**), even when evolutionary generation was taken into account (**Supplementary Figure 7**). Preadaptation might have participated in direct adaptation-induced niche expansion. Further evidence and investigation will be required to draw a solid conclusion.

DISCUSSION

Direct adaptation mediated by experimental evolution (**Figure 2**) was found to trigger correlated adaptation to the environmental gradient, consistently in five *E. coli* strains carrying the reduced genomes of size variation (**Figures 3–5**). This conclusion was drawn from a quantitative evaluation with the newly defined parameter of total ***S***, according to the fundamental niche concept proposed by Hutchinson (1957), which describes a niche without species' interaction and is considered a high-dimensional hypervolume formed by environmental variables. In the present

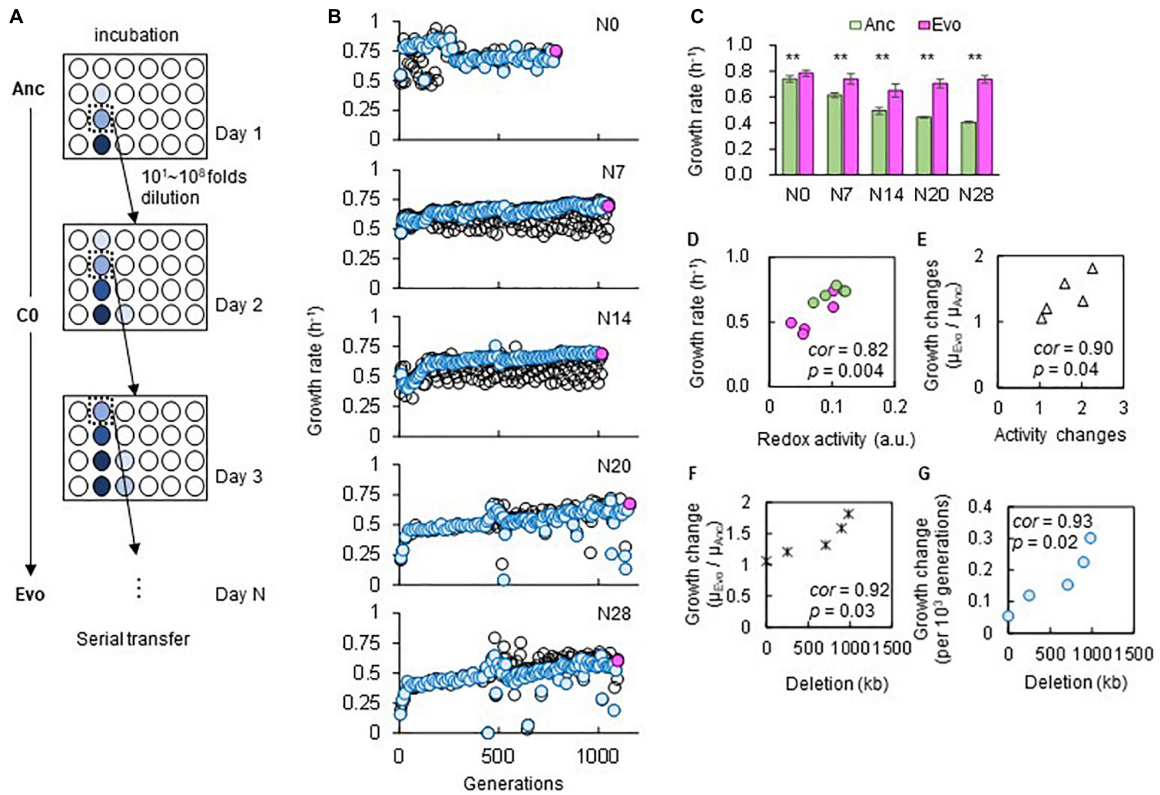


FIGURE 2 | Experimental evolution for adaptation to C0. **(A)** Illustration of experimental evolution. The gradation from light to dark blue indicates the cell concentration from low to high. The *E. coli* cells were cultured in eight wells, and eight 10-fold serial dilutions were prepared. Only one of the eight wells (dilutions) showing growth in the early exponential phase was used for serial transfer, as indicated by dotted boxes. **(B)** Temporal changes in growth rates during the experimental evolution. The five genomes are indicated. The blue circles represent the lineages adopted for the following serial transfer, corresponding to the wells indicated by dotted boxes in **(A)**. The black circles stand for the remaining serial dilutions (i.e., the other seven wells) that were not adopted for serial transfer. The endpoint populations of the evolutionary lineages (blue circles) used for the following analyses are indicated in pink. **(C)** Growth rates in the medium for evolution. Green and pink bars represent the growth rates of Ancs and Evos in C0, respectively. Standard errors of biological replications ($N > 6$) are indicated. Asterisks indicate the statistical significance of the two-tailed Student's *t*-test ($p < 0.01$). **(D,E)** Correlation between the growth rate and cellular redox activity. Green and pink circles represent Ancs and Evos, respectively. The Spearman rank correlation coefficients and statistical significance are indicated. **(F)** Correlation between the changes in growth rate and genome reduction. The lengths of the genomic deletions are plotted against the ratio of the growth rates of Ancs and Evos. The Spearman rank correlation coefficients and statistical significance are indicated. **(G)** Correlation between the changes in growth rate per generation and genome reduction. The Spearman rank correlation coefficients and statistical significance are indicated.

study, the variables were the eight constituents of the culture medium used for evolution (i.e., the chemical niches). The overall fitness equality across the environmental gradient seemed to be homeostatic in a defined habitat, as the total S increased while its variation decreased to comparable levels in five genomes of different reduced lengths (**Figures 5C–E**). A habitat composed of multiple chemical niches might decide the maximal accessibility for evolution. Notably, the homeostasis of niche space was not biased by normalization. As normalization to each single unit was performed individually, the maxima of the overall S could be differentiated among the respective genomes.

It was intriguing to find that correlated adaptation to the environmental gradient resulted from evolution under stable conditions. Since experimental evolution was conducted under stable conditions and bacterial growth was maintained in the exponential phase, neither nutritional starvation nor large environmental fluctuation was assumed to occur. As adaptation to one environment often results in maladaptation to alternative environments (Goddard and Bradford, 2003; Rodriguez-Verdugo et al., 2014; Satterwhite and Cooper, 2015), ecological niche speciation is often explained by adaptive trade-offs (Cooper and Lenski, 2000; Goddard and Bradford, 2003; Sexton et al., 2017; Chavhan et al., 2020). In evolution, environmental homogeneity is considered one of the factors determining trade-offs (Bono et al., 2017) and environmental fluctuation is thought to be crucial for adaptation to a wide range of environments (Wang and Dai, 2019). The reduced genomes evolved in a jack-of-all-trades-and-master-of-all manner, which has been proposed as one of three mechanisms of specialization that is widespread in nature (Remold, 2012); in contrast, the wild-type genome adopted a trade-off mechanism (**Figures 5A–C**), which is generally explained by constraints in phenotypic space (Shoval et al., 2012; Fraebel et al., 2017). In nature, the trade-off strategy might be more common and feasible for costless

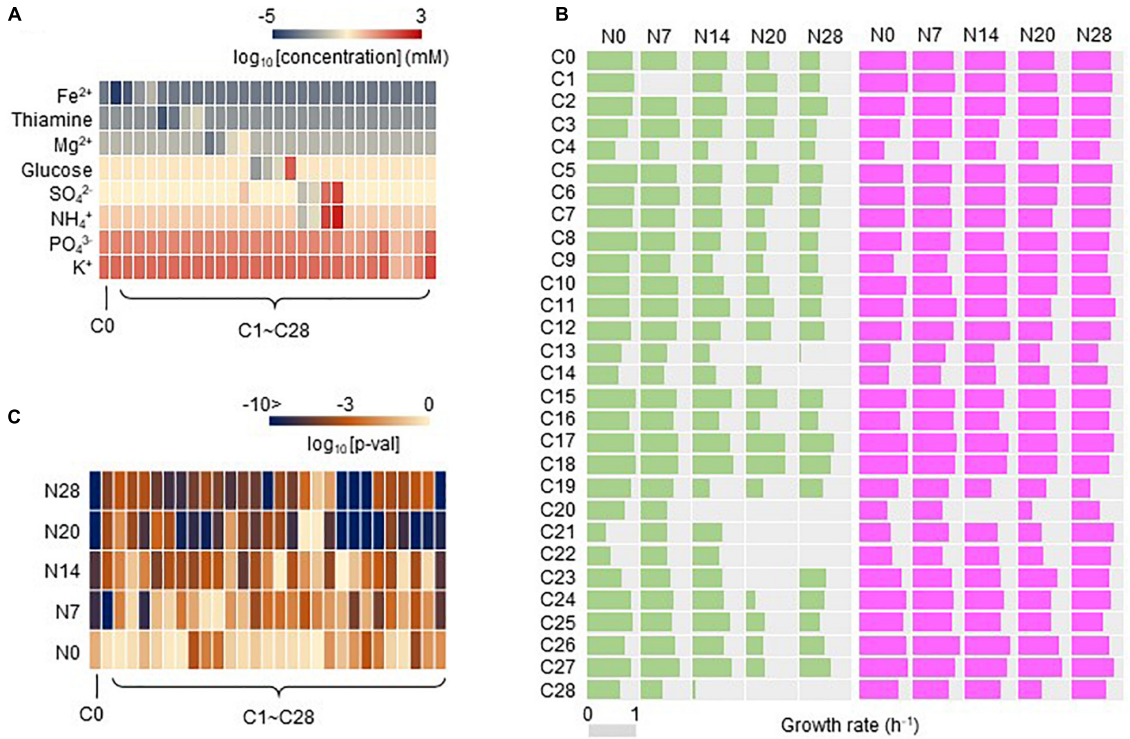


FIGURE 3 | Growth rates evaluated in 29 medium combinations. **(A)** Heatmap of the concentration gradient. The eight chemical components of the media are indicated. The gradation from red to blue indicates the concentration, which is shown on the logarithmic scale, from high to low. C0 is the environment for evolution. **(B)** Growth rates of Ancs and Evos in the environmental gradient. The growth rates evaluated in the 29 medium combinations are shown. Pink and green represent Evos and Ancs, respectively. The gray bar represents the scale of the exact growth rate (h^{-1}). **(C)** Heatmap of statistical significance of the changes in growth rate of Evos and Ancs. The gradation from light orange to dark blue indicates the *p*-value, which is shown on the logarithmic scale, from low to high.

adaptation and niche expansion during eco-evolution (Ferenci, 2016; Farahpour et al., 2018; Bono et al., 2020).

The correlation between ***S*** and genome reduction was significant for glucose, NH_4^+ and SO_4^{2-}, whereas direct adaptation abolished this correlation for NH_4^+ and SO_4^{2-} and changed the correlation from negative to positive for glucose (**Figures 5A,B**). Direct adaptation might directly compensate for the deficiency in using these resources, which seems reasonable because carbon, nitrogen and sulfur are the essential major elements required by living organisms on Earth (Wackett et al., 2004). Such niche specificity might reflect the evolutionary direction of generalists or specialists (Kassen, 2002; Bono et al., 2020). A large omnidirectional expansion of total ***S*** was found in the largely reduced genomes, in comparison to the small directional expansion of total ***S*** in the complete and slightly reduced genomes (**Figure 5C**). This revealed that organisms that experience large genome reduction evolved for generally and less genome reduction evolved for specially, which was an intriguing strategy of genome evolution for niche expansion.

Additionally, genome reduction-dependent features were generally observed. If any functions or mechanisms of the deleted genes were specifically responsible for direct and/or correlated adaptation, the size of the genome reduction would never be correlated with the fitness increase (**Figure 2**) or niche expansion (**Figure 5**). Genome reduction determines correlated adaptation to environmental gradients to some extent. A limitation of the study was lack of replication of experimental evolution, that is, only a single lineage of experimental evolution was applied in the five genomes. Nevertheless, the correlations of genome reduction to adaptation were highly reliable, because a non-correlated relationship will more frequently be acquired by chance. As all the correlations observed in the present study were statistically significant, additional evolutionary lineages would neither change the conclusion nor mask the correlations. Repeated experimental evolution is required to verify the generality of the correlations. Mutations appeared in the adaptive evolution might be fixed occasionally, as they were largely differentiated in gene functions and/or mechanisms (**Supplementary Table 6**). Repeated evolution experiments might result in different mutations in the same genome. This was the reason why the contribution of the mutations to fitness was analyzed from the viewpoint of correlation instead of gene function, which was assumed to be highly differentiated among multiple evolutionary lineages. The present study revealed a quantitative relationship among the genome reduction, adaptation and niche expansion as an experimental demonstration of the linkage between evolution and ecology. Further evaluation of the growth fitness in largely different environments will be highly valuable to clarify the evolutionary trade-offs and/or generality due to genome reduction.

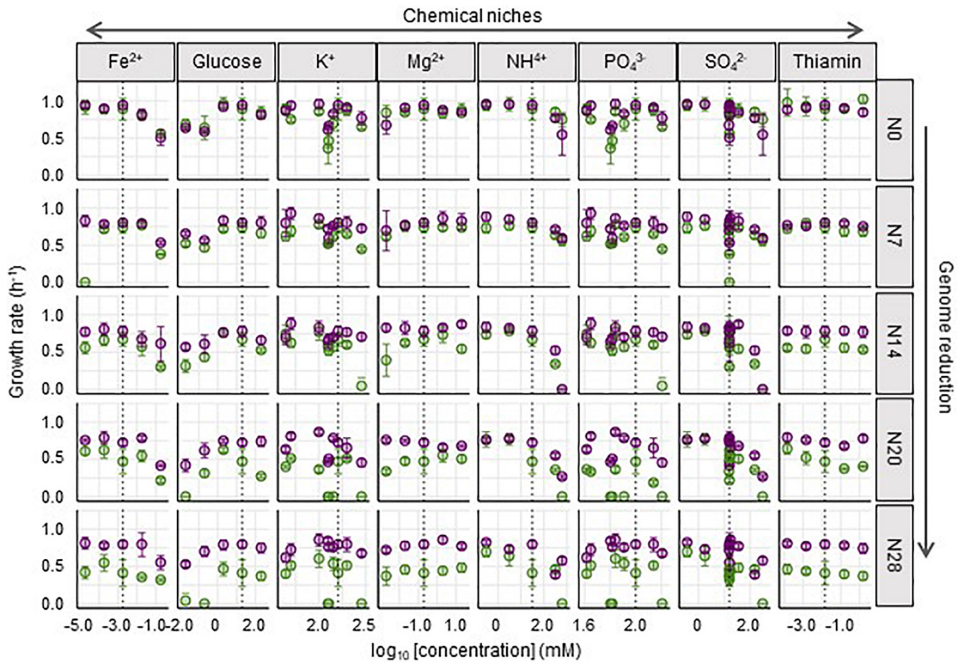


FIGURE 4 | Fitness across the concentration gradient of individual chemical niches. The mean growth rates in the 29 medium combinations are shown. The concentrations of chemical niches are shown on a logarithmic scale. The wild-type and reduced genomes are indicated as N0 and N7∼N28, respectively. Purple and green represent Evos and Ancs, respectively. Standard errors of biological replications ($N > 6$) are indicated.

MATERIALS AND METHODS

Escherichia coli Strains

A total of five *E. coli* strains with either the wild-type or the reduced genome were used, which were selected from the KHK collection (Mizoguchi et al., 2008), an *E. coli* collection of reduced genomes (from National BioResource Project, National Institute of Genetics, Shizuoka, Japan). The wild-type and four reduced genomes were derived from *E. coli* W3110 and were assigned as N0 and N7, 14, 20, 28, respectively (**Supplementary Table 1**), according to previous studies (Kurokawa et al., 2016; Nishimura et al., 2017). Note that genome reductions are additively cumulative. Therefore, the higher numbered strains include all genomic reduction in the lower numbered strains.

Media Combinations

The minimal medium M63, equivalent to C0, was used for the experimental evolution for direct adaptation. Its chemical composition was described in detail previously (Kurokawa et al., 2016; Kurokawa and Ying, 2017). The concentration gradient of the components of the M63 medium was prepared just before the fitness assay by mixing the stock solutions of individual chemical compounds, which resulted in 28 alternative medium combinations (C1∼28). The stock solutions, that is, 1 M glucose, 0.615 M K_2HPO_4, 0.382 M KH_2PO_4, 0.203 M $MgSO_4$, 0.0152 M thiamin/HCl, 0.0018 M $FeSO_4$, and 0.766 M $(NH_4)_2SO_4$, were sterilized using a sterile syringe filter with a 0.22-μm pore size hydrophilic PVDF membrane (Merck, United States). The concentrations of most chemical compounds were altered one-by-one on a logarithmic scale to achieve a wide range of environmental gradients, as described previously (Ashino et al., 2019), which led to a total of 28 combinations. Both the medium used in the evolution (C0) and the alternative medium combinations (C1∼28) were used for the fitness assay. The resultant concentrations of individual components in the ionic form are summarized in **Supplementary Table 2**.

Experimental Evolution

The experimental evolution of the five *E. coli* strains was performed within the early exponential phase by serial transfer, which was performed with 24-well microplates specific for microbe culture (IWAKI, Japan) as previously described (Nishimura et al., 2017). The *E. coli* cells were cultured in eight wells, and eight 10-fold serial dilutions, i.e., 10^1∼10^8, were prepared with fresh medium. The microplates were incubated overnight in a microplate bioshaker (Deep Well Maximizer, Taitec, Japan) at 37°C, with rotation at 500 rpm. Serial transfer was performed at 12- or 24-h intervals, according to the growth rate. Only one of the eight wells (dilutions) showing growth in the early exponential phase (OD_{600} = 0.01–0.1) was selected and diluted into eight wells of a new microplate using eight dilution ratios. The cell culture selected daily for the following serial transfer was mixed with glycerol (15% v/v) and stored at –80°C for future analyses. Serial transfer was repeatedly performed for approximately 50 days. The evolutionary generation was calculated according to the following equation (Eq. 1).

$$gen = log_2(C_i/C_j) \quad (1)$$

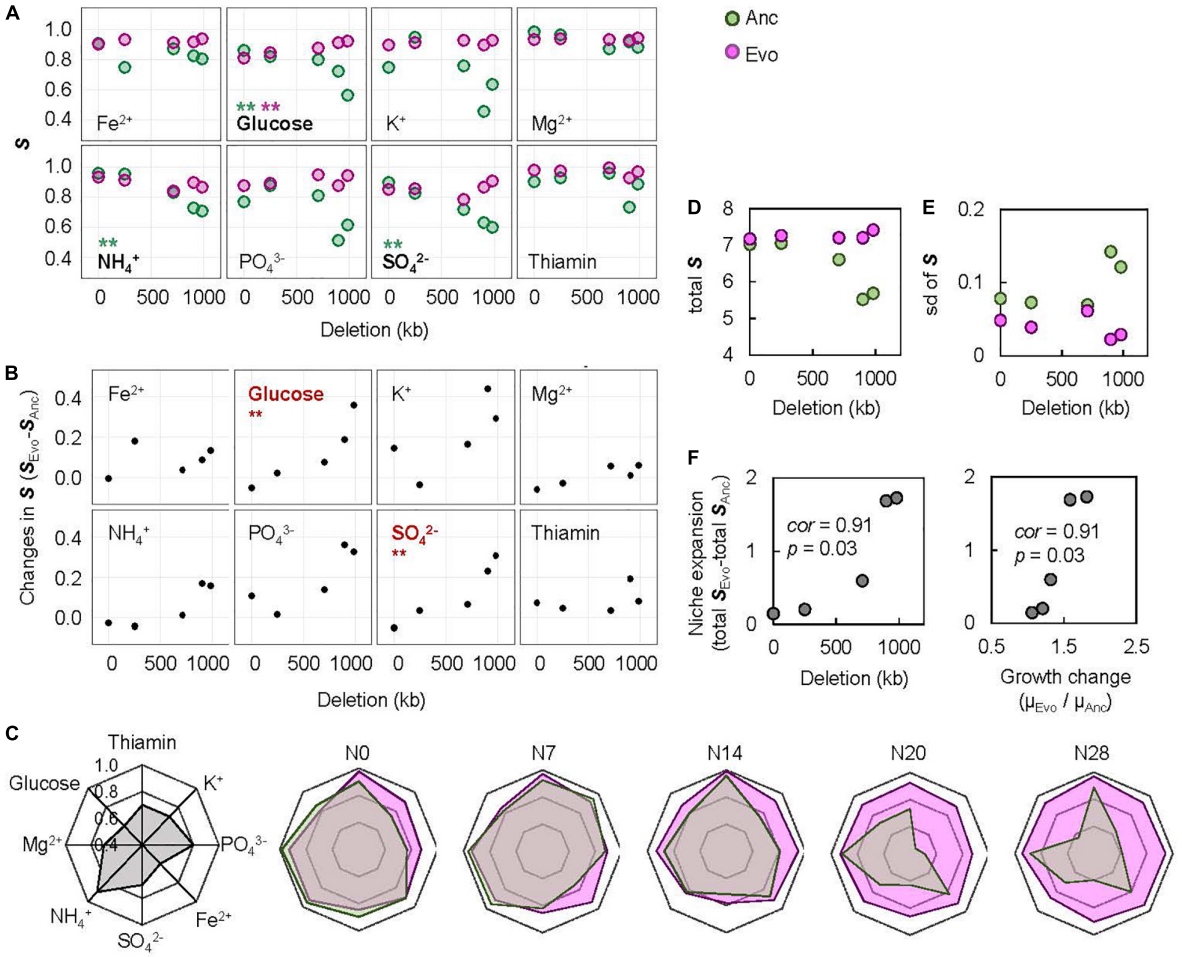


FIGURE 5 | Evaluation of the breadth of adaptation. **(A)** Relationships between the ***S*** and genome reduction. The chemical niches and the statistical significance of the Spearman rank correlation are indicated. Boldfaces associated with asterisks represent statistical significance (**$p < 0.01$). Pink and green represent Evo and Anc, respectively. **(B)** Relationships between the changes in ***S*** and genome reduction. The chemical niches and the statistical significance of the Spearman rank correlation are indicated. Boldfaces in red associated with asterisks represent statistical significance (**$p < 0.01$). **(C)** Radar chart of the eight niche spaces. The eight chemical niches and the scale of ***S*** are illustrated in the monotone radar chart. The five genomes are shown separately. Transparent pink and green represent Evos and Ancs, respectively. **(D)** Sum of the eight ***S***. **(E)** Standard deviation of the eight ***S***. **(F)** Correlation of the changes in total ***S*** to genome reduction and direct adaptation. The Spearman rank correlation coefficients and statistical significance are indicated.

Here, C_i and C_j represent the OD_{600} of the cell culture that was used for serial transfer and the theoretical OD_{600} of the cell culture at the start of incubation. C_j was calculated by dividing the OD_{600} that was used in the last transfer by the dilution rate. To benefit experimental replication, the cell cultures stored for the following assays were dispensed into 20 microtubes in small aliquots (100 μL per tube), which were used once, and the remainder was discarded, as previously described (Kurokawa and Ying, 2017). Growth rate was estimated at every serial transfer according to the following equation (Eq. 2).

$$\mu = LN\left(C_i/C_j\right)/\left(t_i - t_j\right) \quad (2)$$

Here, C_i and C_j are as described above. t_i is the time of serial transfer operation, and t_j is the time of the serial transfer operation immediately before t_i.

Fitness Assay

The fitness was determined as the maximal growth rate, as previously reported (Kurokawa et al., 2016). In brief, the cell culture stocks were diluted 1,000-fold in fresh media (C0, C1∼28) and were subsequently loaded into a 96-well microplate (Costar, United States) in six wells at varied locations. The 96-well microplate was incubated in a plate reader (Epoch2, BioTek) with a rotation rate of 567 rpm at 37°C. The temporal growth of the *E. coli* cells was detected by measuring the absorbance at 600 nm, and readings were obtained at 30-min intervals for 48 h. The maximal growth rate was calculated according to the following equation (Eq. 3).

$$\mu = LN\left(C_{i+1}/C_i\right)/\left(t_{i+1} - t_i\right) \quad (3)$$

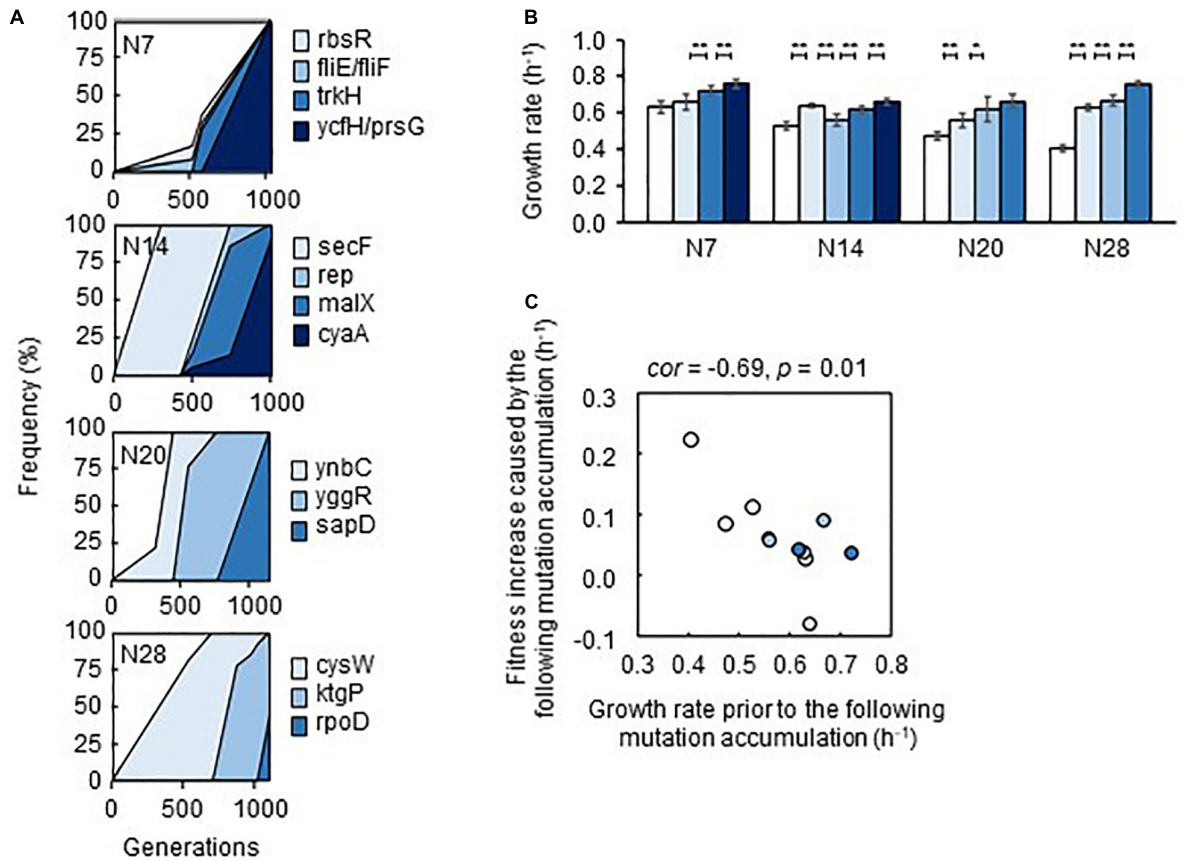


FIGURE 6 | Fitness increase attributed to the mutations. **(A)** Stepwise accumulation of genome mutations. Mutation during the evolution. The temporal changes in the mutations fixed during the evolution are shown. The four reduced genomes and the gene names of mutants are indicated. **(B)** Additive increase in the growth rates of the mutants in the medium for evolution. Gradation from white to dark blue indicates the mutants with respect to those shown in **(A)**. Standard errors of biological replications ($n > 6$) are indicated. Asterisks indicate the statistical significance of the two-tailed Student's *t*-test (*$p < 0.05$; **$p < 0.01$). **(C)** Correlation between mutation accumulation and changes in growth rates. A total of 12 mutants of a single accumulated mutation are shown. The Spearman rank correlation coefficients and statistical significance are indicated.

Here, C_i and C_{i+1} represent the two reads of OD_{600} values at two consecutive time points of t_i and t_{i+1}. The growth fitness was the average of the five continuous growth rates that exhibited the largest mean and the smallest standard deviation during the temporal changes in growth rate, as previously reported (Kurokawa et al., 2016). A total of 2,220 growth curves were acquired, and the corresponding growth rates were calculated for the analysis (**Supplementary Table 3**). Statistical significance of the changes in growth rate mediated by experimental evolution was evaluated by t-student test and the results are summarized in **Supplementary Table 4**.

Redox Activity Assay

A cell culture in the exponential phase of growth (OD_{600} = 0.01 ∼ 0.3) was used for the assay. The cell culture was diluted with fresh medium at 12 dilution ratios from 1.75^0 to 1.75^{11} in a final volume of 2 mL. Every 100 μL of the diluted cell culture was transferred to multiple wells in a 96-well microplate (Costar, United States), in which 20 μL of CellTiter 96® Aqueous One Solution Reagent (Promega) was added. The reduction of the tetrazolium compound in the reagents was measured with a microplate reader (Epoch2, BioTek, United States) by determining the OD_{490} every 2 min for 30 min. The rate of reduction was calculated by linear regression of the temporal changes in OD_{490}, i.e., the slope of the increase in OD_{490} over time (min). The redox activity was determined by dividing the rate of reduction by the OD_{600} of the cell culture. The mean of the multiple measurements ($N = 5$) was used for the analysis.

Evaluation of the Breadth of Adaptation

The fitness dynamics across the concentration gradient of each constituent (ion) were evaluated by curve fitting of a cubic polynomial with the following equation (Eq. 4).

$$\mu(x) = ax^3 + bx^2 + cx + d \quad (4)$$

Here, x and $\mu(x)$ represent the concentration gradient of each constituent and the growth rate under the corresponding conditions, respectively. a, b, c and d are the constants. The area under the regression curve was calculated according to the

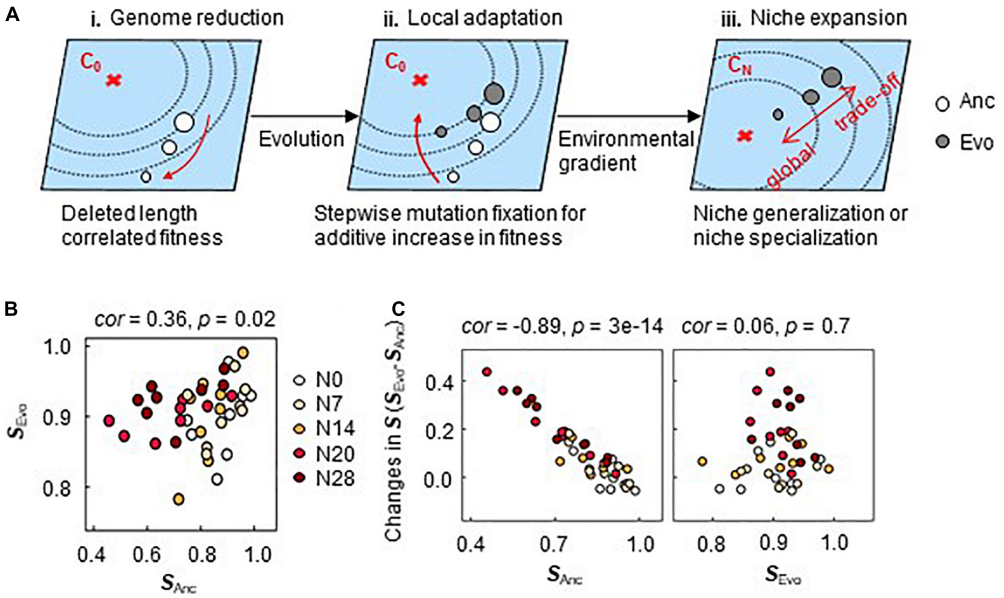


FIGURE 7 | Mechanism for pre-adaptation. **(A)** Illustration of fitness landscape analysis. The fitness landscapes in environment C_0 for the evolution and the alternative environments C_N are shown as the planes in blue. The red cross and the contour lines indicate the fitness peak (maximum) and the fitness gradient, respectively. The open and closed circles indicate the ancestor and evolved genomes, respectively. The size of the circles indicates the genome size. **(B)** Correlated **S** of Ancs and Evos. The color variation from white to dark red represents the five different genomes of N0~N28. The niche spaces of Ancs and Evos are named $\boldsymbol{S_{Anc}}$ and $\boldsymbol{S_{Evo}}$, respectively. **(C)** Relationships of $\boldsymbol{S_{Anc}}$ and $\boldsymbol{S_{Evo}}$ to the changes in **S**. The changes in **S** caused by the evolution are plotted against $\boldsymbol{S_{Anc}}$ and $\boldsymbol{S_{Evo}}$ in the left and right panels, respectively. The Spearman rank correlation coefficients and statistical significance are indicated.

following equation (Eq. 5).

$$Area = \int_{x_{min}}^{x_{max}} ax^3 + bx^2 + cx + d \quad (5)$$

Here, x_{min} and x_{max} represent the minimum and maximum concentrations of each chemical component, respectively. The space (**S**) was evaluated by normalizing both the height and the width of the regression curve with the following equation (Eq. 6).

$$S = Area \times {\mu_{max}}^{-1} \times (x_{max} - x_{min})^{-1} \quad (6)$$

Here, μ_{max} is the maximal growth rate across the concentration gradient. The niche broadness (S_T) of the individual genome was determined as the sum of the niche spaces of the eight chemical components as follows (Eq. 7).

$$S_T = \sum_{i=1}^{n} S_i \quad (7)$$

Here, S_i and n indicate the niche space of each chemical component and the total number of chemical components, respectively.

Genome Resequencing and Mutation Analysis

The stored cell culture was inoculated into 4 mL of fresh M63 medium in a test tube and grown at 37°C with shaking at 200 rpm. Once cell growth reached the stationary phase (OD_{600} > 1.0), rifampicin was added to the culture at a final concentration of 300 μg/mL to stop genome replication initiation. After 3 h of culture with rifampicin, the cells were collected as previously reported (Kishimoto et al., 2010). Genomic DNA was extracted using a Wizard Genomic DNA Purification Kit (Promega, United States) in accordance with the manufacturer's instructions. The sequencing libraries were prepared using the Nextera XT DNA Sample Prep Kit (Illumina), and paired-end sequencing (300 bp × 2) was performed with the Illumina MiSeq platform. The sequencing reads were aligned to the *E. coli* W3110 reference genome (AP009048.1, GenBank), and the genome mutations were analyzed with the Breseq pipeline (version 0.30.1) (Barrick et al., 2014). The statistical data of DNA sequencing and mapping are summarized in **Supplementary Table 5**. The raw data set was deposited in the DDBJ Sequence Read Archive under accession number DRA011629. The fixed mutations (**Supplementary Table 6**) were subsequently analyzed for the temporal order of accumulation during evolution.

Sanger Sequencing and Single-Colony Isolation

The genomic region of approximately 300–600 kb centered on the position of the mutation was amplified by PCR with PrimeSTAR HS DNA Polymerase (TaKaRa Bio, Japan) and the corresponding primers (**Supplementary Table 7**). Amplicons were purified using a MinElute PCR Purification Kit (Qiagen, United States), and Sanger sequencing was conducted by Eurofins Genomics K. K. (Tokyo, Japan). The resulting electropherogram was analyzed using Sequence Scanner Software v2.0 (Thermo Fisher Scientific, United States), and

the ratio of the mutants within the cell population was calculated according to the peak values, as described previously (Kishimoto et al., 2015). Stored cell cultures with an interval of ∼100 generations were analyzed to identify the heterogeneity of the cell population. Single-colony isolation was performed from the heterogeneous population to isolate the homogeneous mutants. The cell culture was spread on LB agar plates, and 10∼30 single colonies per plate were subjected to Sanger sequencing. The colonies of the homogeneous mutant were stored for the fitness assay as described above.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MK and IN performed the experiments. MK and B-WY analyzed the data and drafted the manuscript. B-WY conceived the research and rewrote the manuscript. All authors approved the final manuscript.

FUNDING

This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) (Grant No. 19H03215) and the Grant-in-Aid for Challenging Exploratory Research (Grant No. 21K19815).

ACKNOWLEDGMENTS

We thank NBRP for providing the *E. coli* strains carrying the wild-type and reduced genomes (KHK collection).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.826894/full#supplementary-material

REFERENCES

Alneberg, J., Bennke, C., Beier, S., Bunse, C., Quince, C., Ininbergs, K., et al. (2020). Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun. Biol.* 3:119. doi: 10.1038/s42003-020-0856-x

Andrei, A. S., Salcher, M. M., Mehrshad, M., Rychtecky, P., Znachor, P., and Ghai, R. (2019). Niche-directed evolution modulates genome architecture in freshwater planctomycetes. *ISME J.* 13, 1056–1071. doi: 10.1038/s41396-018-0332-5

Ankrah, N. Y. D., Chouaia, B., and Douglas, A. E. (2018). The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. *mBio* 9:18. doi: 10.1128/mBio.01433-18

Ashino, K., Sugano, K., Amagasa, T., and Ying, B. W. (2019). Predicting the decision making chemicals used for bacterial growth. *Sci. Rep.* 9:7251. doi: 10.1038/s41598-019-43587-8

Bajic, D., Vila, J. C. C., Blount, Z. D., and Sanchez, A. (2018). On the deformability of an empirical fitness landscape by microbial evolution. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11286–11291. doi: 10.1073/pnas.1808485115

Barrick, J. E., and Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nat. Rev. Genet.* 14, 827–839.

Barrick, J. E., Colburn, G., Deatherage, D. E., Traverse, C. C., Strand, M. D., Borges, J. J., et al. (2014). Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genom.* 15:1039. doi: 10.1186/1471-2164-15-1039

Barrick, J. E., Kauth, M. R., Strelioff, C. C., and Lenski, R. E. (2010). *Escherichia coli* rpoB mutants have increased evolvability in proportion to their fitness defects. *Mol. Biol. Evol.* 27, 1338–1347. doi: 10.1093/molbev/msq024

Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* 12, 841–850. doi: 10.1038/nrmicro3331

Baym, M., Lieberman, T. D., Kelsic, E. D., Chait, R., Gross, R., Yelin, I., et al. (2016). Spatiotemporal microbial evolution on antibiotic landscapes. *Science* 353, 1147–1151. doi: 10.1126/science.aag0822

Bono, L. M., Draghi, J. A., and Turner, P. E. (2020). Evolvability costs of niche expansion. *Trends Genet. TIG* 36, 14–23. doi: 10.1016/j.tig.2019.10.003

Bono, L. M., Smith, L. B. Jr., Pfennig, D. W., and Burch, C. L. (2017). The emergence of performance trade-offs during local adaptation: insights from experimental evolution. *Mol. Ecol.* 26, 1720–1733. doi: 10.1111/mec.13979

Chavhan, Y., Malusare, S., and Dey, S. (2020). Larger bacterial populations evolve heavier fitness trade-offs and undergo greater ecological specialization. *Heredity* 124, 726–736. doi: 10.1038/s41437-020-0308-x

Choe, D., Lee JH., Yoo M., Hwang S., Sung BH., Cho S., et al. (2019). Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* 10:935.

Chu, X., Li, S., Wang, S., Luo, D., and Luo, H. (2021). Gene loss through pseudogenization contributes to the ecological diversification of a generalist roseobacter lineage. *ISME J.* 15, 489–502. doi: 10.1038/s41396-020-00790-0

Cooper, V. S., and Lenski, R. E. (2000). The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407, 736–739. doi: 10.1038/35037572

Couce, A., and Tenaillon, O. A. (2015). The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* 6:99. doi: 10.3389/fgene.2015.00099

Cullum, A. J., Bennett, A. F., and Lenski, R. E. (2001). Evolutionary adaptation to temperature. IX. preadaptation to novel stressful environments of *Escherichia coli* adapted to high temperature. *Evol. Int. J. Organic Evol.* 55, 2194–2202. doi: 10.1111/j.0014-3820.2001.tb00735.x

Daubin, V., and Szöllősi, G. J. (2016). Horizontal gene transfer and the history of life. *Cold Spring Harbor Perspect. Biol.* 8:a018036.

de Visser, J. A., and Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* 15, 480–490. doi: 10.1038/nrg3744

Farahpour, F., Saeedghalati, M., Brauer, V. S., and Hoffmann, D. (2018). Trade-off shapes diversity in eco-evolutionary dynamics. *Elife* 7:e36273. doi: 10.7554/eLife.36273

Ferenci, T. (2016). Trade-off mechanisms shaping the diversity of bacteria. *Trends Microbiol.* 24, 209–223. doi: 10.1016/j.tim.2015.11.009

Fraebel, D. T., Mickalide, H., Schnitkey, D., Merritt, J., Kuhlman, T. E., and Kuehn, S. (2017). Environment determines evolutionary trajectory in a constrained phenotypic space. *Elife* 6:e24669. doi: 10.7554/eLife.24669

Fragata, I., Blanckaert, A., Dias Louro, M. A., Liberles, D. A., and Bank, C. (2019). Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* 34, 69–82. doi: 10.1016/j.tree.2018.10.009

Getz, E. W., Tithi, S. S., Zhang, L., and Aylward, F. O. (2018). Parallel evolution of genome streamlining and cellular bioenergetics across the marine radiation of a bacterial phylum. *mBio* 9:5. doi: 10.1128/mBio.01089-18

Goddard, M. R., and Bradford, M. A. (2003). The adaptive response of a natural microbial population to carbon- and nitrogen-limitation. *Ecol. Lett.* 6, 594–598. doi: 10.1046/j.1461-0248.2003.00478.x

Graham, E. D., and Tully, B. J. (2021). Marine dadabacteria exhibit genome streamlining and phototrophy-driven niche partitioning. *ISME J.* 15, 1248–1256. doi: 10.1038/s41396-020-00834-5

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia Quant. Biol.* 22, 415–427.

Jasmin, J. N., and Zeyl, C. (2013). Evolution of pleiotropic costs in experimental populations. *J. Evol. Biol.* 26, 1363–1369. doi: 10.1111/jeb.12144

Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., et al. (2016). Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Mol. Biol. Evol.* 33, 1257–1269. doi: 10.1093/molbev/msw009

Kassen, R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* 15, 173–190. doi: 10.1046/j.1420-9101.2002.00377.x

Kato, J., and Hashimoto, M. (2007). Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3:132. doi: 10.1038/msb4100174

Kawecki, T. J., and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecol. Lett.* 7, 1225–1241. doi: 10.1111/j.1461-0248.2004.00684.x

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M. C. (2012). Experimental evolution. *Trends Ecol. Evol.* 27, 547–560.

Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618.

Kishimoto, T., Iijima, L., Tatsumi, M., Ono, N., Oyake, A., Hashimoto, T., et al. (2010). Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. *PLoS Genet.* 6:e1001164. doi: 10.1371/journal.pgen.1001164

Kishimoto, T., Ying, B. W., Tsuru, S., Iijima, L., Suzuki, S., Hashimoto, T., et al. (2015). Molecular clock of neutral mutations in a fitness-increasing evolutionary process. *PLoS Genet.* 11:e1005392. doi: 10.1371/journal.pgen.1005392

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

Kuo, C. H., and Ochman, H. (2009). Deletional bias across the three domains of life. *Geno. Biol. Evol.* 1, 145–152. doi: 10.1093/gbe/evp016

Kurokawa, M., and Ying, B. W. (2017). Precise, high-throughput analysis of bacterial growth. *J. Vis. Exp.* 127:56197. doi: 10.3791/56197

Kurokawa, M., Seno, S., Matsuda, H., and Ying, B. W. (2016). Correlation between genome reduction and bacterial growth. *DNA Res. Int. J. Rapid Publi. Rep. Genes Genom.* 23, 517–525. doi: 10.1093/dnares/dsw035

Lynch, M., and Gabriel, W. (1987). Environmental tolerance. *Am. Nat.* 129, 283–303. doi: 10.1201/b10519-18

Maistrenko, O. M., Mende, D. R., Luetge, M., Schmidt, T. S. B., and Li, S. S. (2020). Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* 14, 1247–1259. doi: 10.1038/s41396-020-0600-z

Martin, G., and Lenormand, T. (2015). The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evol. Int. J. Organic Evol.* 69, 1433–1447. doi: 10.1111/evo.12671

Mizoguchi, H., Sawano, Y., Kato, J., and Mori, H. (2008). Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res. Int. J. Rapid Publi. Rep. Genes Genom.* 15, 277–284. doi: 10.1093/dnares/dsn019

Nahum, J. R., Godfrey-Smith, P., Harding, B. N., Marcus, J. H., Carlson-Stevermer, J., and Kerr, B. (2015). A tortoise-hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7530–7535. doi: 10.1073/pnas.1410631112

Nishimura, I., Kurokawa, M., Liu, L., and Ying, B. W. (2017). Coordinated changes in mutation and growth rates induced by genome reduction. *mBio* 8:4. doi: 10.1128/mBio.00676-17

Posfai, G., Plunkett, G. III, 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

Rees-Garbutt, J., Chalkley, O., Landon, S., Purcell, O., Marucci, L., and Grierson, C. (2020). Designing minimal genomes using whole-cell models. *Nat. Commun.* 11:836.

Remold, S. (2012). Understanding specialism when the jack of all trades can be the master of all. *Proc. R. Soc. B Biol. Sci.* 279, 4861–4869. doi: 10.1098/rspb.2012.1990

Rodriguez-Verdugo, A., Carrillo-Cisneros, D., Gonzalez-Gonzalez, A., Gaut, B. S., and Bennett, A. F. (2014). Different tradeoffs result from alternate genetic adaptations to a common environment. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12121–12126. doi: 10.1073/pnas.1406886111

Salcher, M. M., Schaefle, D., Kaspar, M., Neuenschwander, S. M., and Ghai, R. (2019). Evolution in action: habitat transition from sediment to the pelagial leads to genome streamlining in methylophilaceae. *ISME J.* 13, 2764–2777. doi: 10.1038/s41396-019-0471-3

Satterwhite, R. S., and Cooper, T. F. (2015). Constraints on adaptation of *Escherichia coli* to mixed-resource environments increase over time. *Evol. Int. J. Organic Evol.* 69, 2067–2078. doi: 10.1111/evo.12710

Schick, A., Bailey, S. F., and Kassen, R. (2015). Evolution of fitness trade-offs in locally adapted populations of *Pseudomonas* fluorescens. *Am. Natur.* 186, S48–S59. doi: 10.1086/682932

Sexton, J. P., Montiel, J., Shay, J. E., Stephens, M. R., and Slatyer, R. A. (2017). Evolution of ecological niche breadth. *Annu Rev. Ecol. Evol.* 48, 183–206.

Shoval, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., et al. (2012). Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. *Science* 336, 1157–1160.

Szamecz, B., Boross G., Kalapis D., Kovács K., Fekete G., Farkas Z., et al. (2014). The genomic landscape of compensatory evolution. *PLoS Biol.* 12:e1001935. doi: 10.1371/journal.pbio.1001935

Tenaillon, O. (2014). The utility of Fisher's geometric model in evolutionary genetics. *Annu Rev. Ecol. Evol. Syst.* 45, 179–201. doi: 10.1146/annurev-ecolsys-120213-091846

Wackett, L. P., Dodge, A. G., and Ellis, L. B. (2004). Microbial genomics and the periodic table. *Appl. Environ. Microbiol.* 70, 647–655. doi: 10.1128/AEM.70.2.647-655.2004

Wang, S., and Dai, L. (2019). Evolving generalists in switching rugged landscapes. *PLoS Comput. Biol.* 15:e1007320. doi: 10.1371/journal.pcbi.1007320

Xavier, J. C., Patil, K. R., and Rocha, I. (2014). Systems biology perspectives on minimal and simpler cells. *Microbiol. Mol. Biol. Rev.* 78, 487–509. doi: 10.1128/MMBR.00050-13

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fmicb.2022.838267

Long-Read Sequencing Reveals Genetic Adaptation of *Bartonella* Adhesin A Among Different *Bartonella henselae* Isolates

*Arno Thibau[1], Katharina Hipp[2], Diana J. Vaca[1], Sounak Chowdhury[3], Johan Malmström[3], Athanasios Saragliadis[4], Wibke Ballhorn[1], Dirk Linke[4] and Volkhard A. J. Kempf[1]**

[1] Institute for Medical Microbiology and Infection Control, University Hospital, Goethe University, Frankfurt am Main, Germany, [2] Electron Microscopy Facility, Max Planck Institute for Developmental Biology, Tübingen, Germany, [3] Division of Infection Medicine, Department of Clinical Sciences, Lund University, Lund, Sweden, [4] Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, Oslo, Norway

OPEN ACCESS

Edited by:
Daniel Yero,
Universidad Autónoma de Barcelona, Spain

Reviewed by:
Hongkuan Deng,
Shandong University of Technology, China
Burt Anderson,
University of South Florida, United States
Michael F. Minnick,
University of Montana, United States

***Correspondence:**
Volkhard A. J. Kempf
volkhard.kempf@kgu.de

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

Received: 17 December 2021
Accepted: 17 January 2022
Published: 07 February 2022

Citation:
Thibau A, Hipp K, Vaca DJ, Chowdhury S, Malmström J, Saragliadis A, Ballhorn W, Linke D and Kempf VAJ (2022) Long-Read Sequencing Reveals Genetic Adaptation of Bartonella Adhesin A Among Different Bartonella henselae Isolates. Front. Microbiol. 13:838267. doi: 10.3389/fmicb.2022.838267

Bartonella henselae is the causative agent of cat scratch disease and other clinical entities such as endocarditis and bacillary angiomatosis. The life cycle of this pathogen, with alternating host conditions, drives evolutionary and host-specific adaptations. Human, feline, and laboratory adapted *B. henselae* isolates often display genomic and phenotypic differences that are related to the expression of outer membrane proteins, for example the *Bartonella* adhesin A (BadA). This modularly-structured trimeric autotransporter adhesin is a major virulence factor of *B. henselae* and is crucial for the initial binding to the host via the extracellular matrix proteins fibronectin and collagen. By using next-generation long-read sequencing we demonstrate a conserved genome among eight *B. henselae* isolates and identify a variable genomic *badA* island with a diversified and highly repetitive *badA* gene flanked by *badA* pseudogenes. Two of the eight tested *B. henselae* strains lack BadA expression because of frameshift mutations. We suggest that active recombination mechanisms, possibly via phase variation (i.e., slipped-strand mispairing and site-specific recombination) within the repetitive *badA* island facilitate reshuffling of homologous domain arrays. The resulting variations among the different BadA proteins might contribute to host immune evasion and enhance long-term and efficient colonisation in the differing host environments. Considering the role of BadA as a key virulence factor, it remains important to check consistently and regularly for BadA surface expression during experimental infection procedures.

Keywords: trimeric autotransporters adhesin, genetic variability, PacBio SMRT sequencing, host matrix-pathogen interaction, *Bartonella* adhesin A

Abbreviations: ANI, average nucleotide identity; BadA, *Bartonella* adhesin A; BALI, *Bartonella* liquid (medium); BSA, bovine serum albumine; CBA, Columbia blood agar (plates); CCS, circular consensus sequences (reads); Col-I, collagen-I; DAPI, 4′,6-diamidino-2-phenylindole; DMF, dimethylformamide; ECM, extracellular matrix; Fn, fibronectin; FS, freeze substitution; HMW, high molecular weight; HPF, high-pressure freezing; MS, mass spectrometry; o/n, overnight; PLT, progressive lowering of temperature; RT, room temperature; SMRT, single-molecule real-time; TAA, trimeric autotransporter adhesin; TMB, 3,3′,5,5′-tetramethylbenzidine (liquid substrate).

INTRODUCTION

Bartonella henselae is a slow-growing bacterium causing cat scratch disease, a self-limiting zoonotic disease characterised by localised lymphadenopathy or "culture-negative" endocarditis. Infection of immunocompromised patients might result in vasculoproliferative disorders, e.g., bacillary angiomatosis (Relman et al., 1990; Lamps and Scott, 2004). Cats serve as the major reservoir host for *B. henselae* and their infection mostly manifests as asymptomatic bacteraemia. Transmission among cats is predominantly mediated by the cat flea, *Ctenocephalides felis* (Chomel et al., 1996).

The genus *Bartonella* currently consists of more than 40 identified species (Breitschwerdt, 2017; Okaro et al., 2017) and includes both zoonotic and human pathogens with a wide array of mammals as reservoir hosts (Buffet et al., 2013; Kešnerová et al., 2016). *Bartonella* species are haematophagous-arthropod-borne, facultative intracellular α-proteobacteria, and are characterised by their "stealthy" course of infection where host-specific adaptation is essential for their survival (Engel et al., 2011; Harms and Dehio, 2012). As such, long-lasting infections of reservoir hosts are commonly asymptomatic, while incidental host infections often show a clinically apparent course of infection. Homologous recombination (e.g., phase variation) and horizontal gene transfer (HGT) mediate the origination of different substrains within one species, and simultaneously maintain genome integrity (Kyme et al., 2003; Segers et al., 2017; Wagner and Dehio, 2019). Reductive genome evolution is commonly observed within the genus and is concordant with the overall intracellular lifestyle and vector-borne transmission (Ettema and Andersson, 2009).

Consequently, *B. henselae* isolates demonstrate variable genomic and phenotypic differences, presumably driven by adaptation to varying host conditions and caused by frequent recombination events (Lindroos et al., 2006). So far, two *B. henselae* genotypes were proposed based on the 16rRNA gene sequence and are represented by either strain ATCC49882^T Houston-I (genotype I) (Regnery et al., 1992) or strain Marseille (genotype II) (Drancourt et al., 1996). Differing correlations of these genotypes regarding their infection strategy of feline and human endothelial cell lines have been described (Berrich et al., 2011; Chang et al., 2011; Huwyler et al., 2017). Phenotypic differences are among others related to the expression of *Bartonella* adhesin A (BadA) and the VirB/D4 type IV secretion system (Kyme et al., 2003; Lu et al., 2013).

Bartonella henselae is characterised by its enormous surface-expressed adhesin BadA which is important for efficient adherence to extracellular matrix (ECM) proteins (e.g., fibronectin and collagen) and host cells, and for angiogenic reprogramming (Riess et al., 2004; Müller et al., 2011; Kaiser et al., 2012). BadA is a trimeric autotransporter adhesin (TAA) (synonyms: non-fimbrial adhesin, oligomeric coiled-coil adhesin, type Vc secretion system) and follows the common TAA-architecture consisting of an N-terminal head, a repetitive and long neck/stalk region, and a more conserved C-terminal membrane anchor (Hoiczyk, 2000; Szczesny and Lupas, 2008; Leo et al., 2012). The long neck/stalk region is composed of several domains which are defined by a distinctive neck sequence functioning as a connector from bulkier β-strands to slimmer α-helixes (Bassler et al., 2015). The modular and repetitive composition of BadA (and other TAAs) suggests the frequent occurrence of recombination, likely as an evolutionary adaptive process in the light of alternating host conditions. Differences in the repetitive neck/stalk region leading to variations in the size of *badA* among several *B. henselae* strains were previously demonstrated (Riess et al., 2007). However, it was not possible to correctly sequence this genomic region because of its highly repetitive nature. Compared to short-read sequencing technologies, long-read sequencing facilitates to differentiate close variants and to cover highly repetitive stretches without major assembly problems (Tørresen et al., 2019).

In this study we report on long-read sequenced whole genomes of eight *B. henselae* isolates, discuss their genomic organisation with a special emphasis on the highly variable *badA* coding region (i.e., *badA* island), and check for BadA expression and functional binding to ECM proteins. These data suggest an adaptive evolution of the *badA* island and confirm the importance of BadA as a virulence factor.

MATERIALS AND METHODS

Bacterial Strains and Culture Conditions

All *B. henselae* strains used in this study are listed in **Table 1**. Bacteria were grown in *Bartonella* liquid (BALI) medium (Riess et al., 2008) supplemented with 10% sterile fetal calf serum (FCS) for three days in a humidified atmosphere at 37°C and 5% CO_2 while gently shaking (120 RPM). Alternatively, *B. henselae* strains were cultured on Columbia blood agar (CBA) plates with 5% sheep blood (Becton-Dickinson) for either 4 days to obtain fully grown plates, or for 14 days to obtain single colonies, both in a humidified atmosphere at 37°C and 5% CO_2. Both growth conditions ensured surface expression of BadA (proven by immunofluorescence; data not shown). Competent *Escherichia coli* DH5α (NEB), used for cloning and plasmid amplification, were grown overnight (o/n) at 37°C either in shaking (180 RPM) Luria/Miller (LB) broth or on LB agar plates (Carl Roth).

As selection marker, kanamycin (KAN; MP Biomedicals) was used at a final concentration of 30 μg/ml (*B. henselae* Marseille ΔBadA-T) and 50 μg/ml (*E. coli* DH5α). Bacteria were collected by centrifugation at 5,000 × *g* for 10 min at 4°C, unless noted otherwise. Bacterial cryostocks were prepared in LB medium with 20% glycerol and stored at −80°C.

Isolation of Genomic DNA and Whole Genome Sequencing

Bartonella henselae strains grown in BALI medium (isolated from a single colony to avoid a mixture of genomically rearranged subclones) were washed three times with phosphate-buffered saline (PBS; pH 7.2). High molecular weight (HMW) DNA was isolated using the MagAttract HMW DNA kit (Qiagen) and subsequently sheared to ca. 10–12 kb fragments with g-TUBEs (Covaris). The sequencing library was prepared following the Pacific Biosciences (PacBio) protocol for single-molecule real-time (SMRT)bell^{TM} libraries using PacBio®Barcoded Adapters for multiplex SMR®Sequencing. Library samples were size

TABLE 1 | Overview of *B. henselae* strains used in this study[a].

B. henselae strain	Alternative designation(s)	BadA expression	Complete genome sequence	Specifications	Source
Marseille	URLLY-8	Yes	This study	Isolate from a patient diagnosed with cat scratch disease (Marseille, France)	Drancourt et al., 1996
Marseille BadA-transposon mutant	ΔBadA-T	No	–	*B. henselae* Marseille with a TN < KAN-2 > transposon integrated in *badA*	Riess et al., 2003; Riess et al., 2004
Marseille BadA-deletion mutant	ΔBadA-D	No	–	*B. henselae* Marseille with *badA* deleted via homologous recombination	This study
ATCC49882^T var-1	RSE247; CHDE101; Houston-I variant-1	No	This study and NZ_CP020742.1	Spontaneous laboratory streptomycin-resistant variant of *B. henselae* ATCC49882^T Houston-I (isolate from a febrile patient infected with human immunodeficiency virus, Houston, United States)	Regnery et al., 1992; Schmid et al., 2004
ATCC49882^T var-2	Houston-I variant-2	Yes	This study	Laboratory isolate (1996); variant of ATCC49882^T Houston-I	Riess et al., 2007; Lu et al., 2013
Berlin-I	–	No	This study	Isolate from a skin biopsy specimen from a patient diagnosed with bacillary angiomatosis (Berlin, Germany)	Arvand et al., 1998
G-5436	Houston-I; Zürich	Yes	This study	Human isolate, Centers for Disease Control and Prevention (Atlanta, United States); Possible derivative of *B. henselae* ATCC49882^T Houston-I	Regnery et al., 1992; Zbinden et al., 1995; Zbinden et al., 1997
88-64 Oklahoma	–	Yes	This study	Blood isolate from a patient diagnosed with HIV (Oklahoma City, United States)	Welch et al., 1992
FR96/BK38	Type I	Yes	This study	Blood isolate from domestic cat (Freiburg, Germany)	Sander et al., 1997
FR96/BK3	Type II	Yes	This study	Blood isolate from domestic cat (Freiburg, Germany)	Sander et al., 1997

[a]*Passage number of all strains in Frankfurt am Main was <10. Exact passage number before arriving is unknown.*

selected using 0.45× AMPure PB beads and sequenced in a single run (i.e., movie and pre-extension time of 20 and 4 h, respectively) on a PacBio Sequel instrument using v3.0 sequencing chemistry, Sequel polymerase v3.0, and an SMRT cell v3 LR tray.

Reads were demultiplexed using Barcoding pipeline on SMRT Link Analysis Services (v5.1.0.26412 and GUI v5.1.0.26411) with a barcode score of ≥26. *De novo* genome assembly was performed using the HGAP 4 pipeline via SMRT Link Analysis Services (v6.0.0.47841, and GUI v6.0.0.47836) with an expected genome size of 2 Mbp, resulting in single contigs (**Figure 1**). Circular consensus sequences (CCS) reads were computed for the demultiplexed dataset with ≤1 as a number of passes and ≤0.9 as predicted accuracy (**Supplementary Table 1**).

Genomic Analysis and Bioinformatic Tools

Genomes were annotated via both the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation pipeline (Tatusova et al., 2016) and the RASTtk pipeline (Aziz et al., 2008; Brettin et al., 2015). Gene annotation in the *badA* island and flanking up- and downstream regions were manually revised (**Figure 2**). Correct mapping, assembly, and coverage quality of specific genomic regions were checked *in silico* using CCS reads (≥Q20) uploaded in Minimap2 software (Li, 2018) as a Geneious Prime 2020.0.5 plug-in.

Complete multiple alignment of the *B. henselae* conserved genomic sequences was performed using progressiveMAUVE software (Darling et al., 2004, 2010). ProgressiveMAUVE identifies so-called locally collinear blocks among the various strains that prove to be internally free from genome rearrangements (**Figure 1**). Genome comparison as per average nucleotide identity (ANI) based on MUMmer (%) was done with the JSpecies Web Server (Kim et al., 2014; Richter et al., 2016; **Table 2**). Furthermore, potential prophage sequences within the studied *B. henselae* genomes were predicted using Phage Search Tool Enhanced Release (PHASTER) software (Zhou et al., 2011; Arndt et al., 2016). Remote homologues of certain unspecified open reading frames (ORF) were identified in the PDB and Pfam-A databases using the HHpred software (Söding, 2005; Hildebrand et al., 2009; Zimmermann et al., 2018; Steinegger et al., 2019). Further general sequence analyses were consistently performed via SnapGene software (Insightful Science). All mentioned systems and software were run on default parameters.

Generation of a *Bartonella henselae* Marseille *badA* In-Frame Deletion Mutant

A markerless *badA* deletion mutant in *B. henselae* Marseille was generated via a two-step selection process as previously described (Stahl et al., 2015; Weidensdorfer et al., 2016). Primers used in this study are listed in **Supplementary Table 2**.

Vector pBIISK_*sacB*/*kanR*_UpBadA_DownBadA was constructed by ligating two 1-kb flanking regions from up- and downstream the *badA* gene (Marseille), into the linearised pBIISK_*sacB*/*kanR* plasmid directly downstream of *kanR* via Gibson Assembly®(NEB; Gibson et al., 2009). The first fragment contains a 1-kb upstream non-coding region, as well as 30-bp of 5′-*badA* (amplified using primers FrUp_Fw and FrUp_Rv). The second fragment contains 30-bp of 3′-*badA*, as well as a 1-kb downstream non-coding region (amplified using the primers

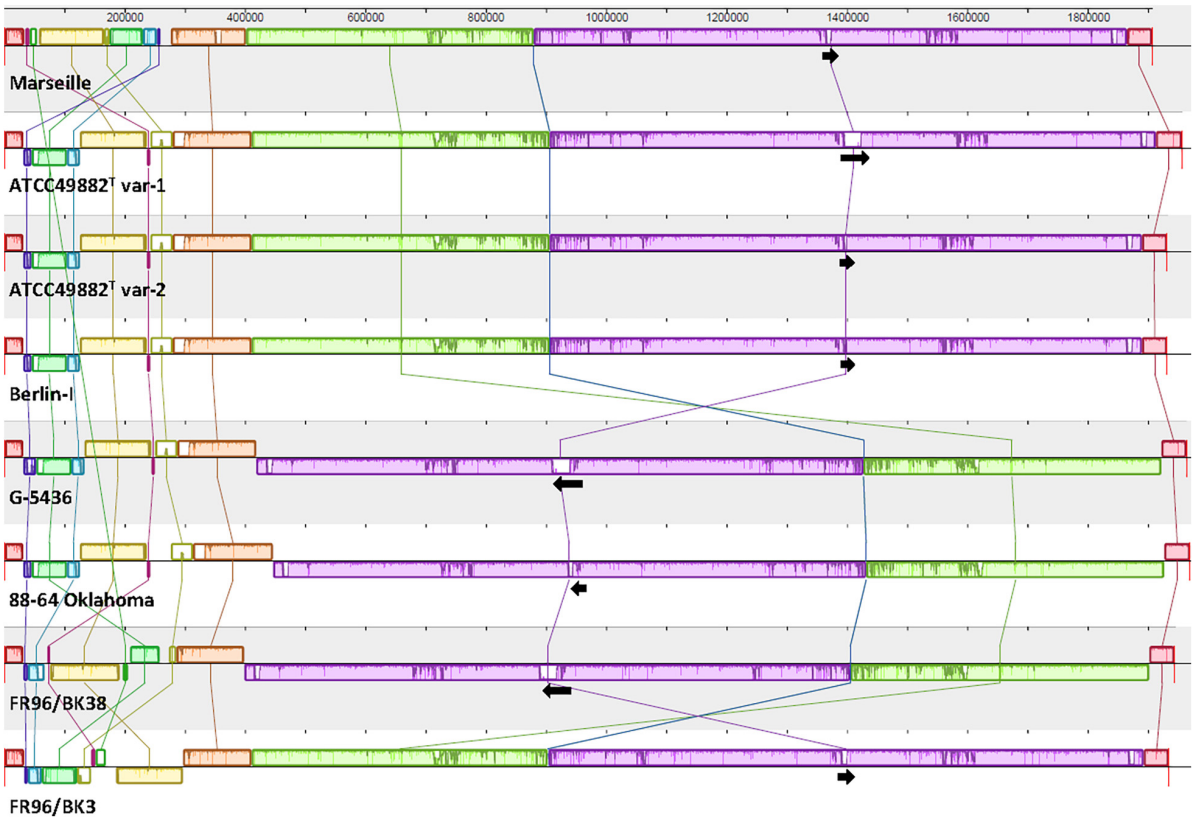


FIGURE 1 | Comparative genome alignment of eight *B. henselae* strains. A genome alignment, visualised via progressiveMAUVE, displays a conserved *B. henselae* genome sequence with few inversions, mostly located in the first ca. 300,000 bp. Genomes are shown as horizontal panels following a black line. Coloured blocks and vertical lines depict localised collinear regions of the genome sequence that align to part of another genome, are homologous, and are internally free from genomic rearrangements. Blocks above or below the horizontal line are in the same or reverse complement orientation compared to the reference genome of *B. henselae* Marseille, respectively. Regions outside blocks lack detectable homology. Inside each block, a similarity profile of the genome sequence is drawn. The height of this profile corresponds to the average level of conservation in that particular region. Complete white areas are not aligned and probably contain sequence elements specific to that strain genome. Black arrows indicate the length, orientation and position of the *badA* island within each genome sequence. The upper scale gives sequence coordinates.

FrDown_Fw and FrDown_Rv). Plasmid pBIISK_*sacB*/*kanR* was linearised using primers pBIISK_Fw and pBIISK_Rv.

pBIISK_*sacB*/*kanR*_UpBadA_DownBadA was propagated in heat-shock transformed *E. coli* DH5α and subsequently electroporated in *B. henselae* Marseille using a Gene Pulser II electroporator (Bio-Rad) as previously described (Riess et al., 2003). Briefly, ca. 4×10^8 electro-competent *B. henselae* Marseille cells were electroporated with 10 μg of purified plasmid DNA, including 1 μl of TypeOne^{TM} Restriction Inhibitor (Lucigen), and immediately incubated in 1 ml recovery broth for 4 h in a humidified atmosphere at 37°C with 5% CO_2 while gently shaking (120 RPM).

Transformed bacteria were subsequently incubated for positive selection on KAN-supplemented CBA plates and resulting clones were checked for integration of pBIISK_*sacB*/*kanR*_UpBadA_DownBadA in the genomic DNA by colony PCR using various primer combinations (pBIISK_Fw, pBIISK_Rv, IntegrationA_Fw, IntegrationA_Rv, IntegrationB_Fw, and IntegrationB_Rv). Clones with a correct insertion were transferred onto CBA-plates supplemented with 10% sucrose to facilitate and select for segregation of the integrated suicide vector together with the *badA* gene. Grown colonies were further identified to check for proper vector loss on CBA plates with and without KAN. Finally, correct *badA* deletion was verified by colony PCR, Sanger sequencing (primers in **Supplementary Table 2**), and Western blotting using rabbit anti-BadA IgG antibodies (see below) using *B. henselae* Marseille as positive control.

Transmission Electron Microscopy

Bartonella henselae strains were grown in BALI medium starting from single colonies and fixed with 4% paraformaldehyde and 2.5% glutaraldehyde (both Electron Microscopy Sciences) in 0.1 M phosphate buffer (pH 7.4) for 90 min at RT. Fixed samples were stored at 4°C until processed either by progressive lowering of temperature (PLT) in dimethylformamide (DMF) and embedding in Lowicryl K4M (protocol adapted from Bayer et al., 1985), or by high-pressure freezing (HPF), freeze substitution (FS) and Epon embedding. Average length of expressed BadA was determined using 24 (for strain FR96/BK38) to 65 (for strain Marseille) bacterial cell images.

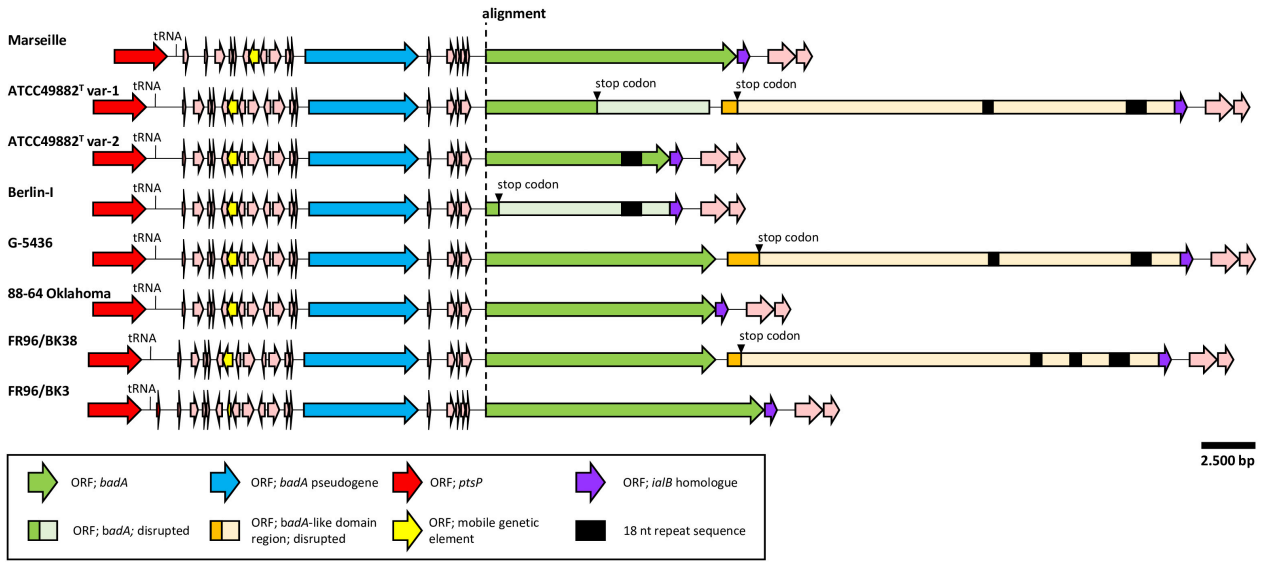


FIGURE 2 | Comparative genetic organisation of the chromosomal *badA* island region. All regions are shown in the same orientation. The conserved upstream ORF (blue) is a *badA* pseudogene containing the typical TAA C-terminal anchor domain. A higher sequence variability is found within the *badA* gene (green), where a premature stop codon is observed in strains ATCC49882^T var-1 and Berlin-I. Strains ATCC49882^T var-1, G-5436, and FR96/BK38 share an exceptionally long ORF (orange), named the *badA*-like domain region that is located downstream of *badA*. These three ORFs also contain the typical TAA C-terminal anchor domain. Additional small ORFs are found between the upstream *badA* pseudogene and *badA* as well as other ORFs flanking the *badA* island (pink). In five displayed *badA* island sequences, two or three unique repeat regions (black rectangle) in either *badA* or in the downstream BadA-like domain region are detected.

For PLT in DMF, fixed bacteria were washed twice in phosphate buffer (1,500 × *g*) and embedded in 12% melted (37°C) gelatine (Merck). Solidified samples were sliced into 1 mm^3 cubes and fixed in 1% glutaraldehyde for 5 min at 4°C. Samples were dehydrated by gradually increasing the DMF concentration from 30% DMF (in H_2O) for 30 min at 0°C to 100% DMF for 1 h at −35°C. Lowicryl K4M was infiltrated at −35°C and polymerised by UV. For HPF/FS, fixed bacteria were successively cryofixed in cellulose capillaries in planchettes filled with 1-hexadecene in a high-pressure freezer (Compact 03, Wohlwend), freeze-substituted in 2% osmium tetroxide/0.4% uranyl acetate in acetone, and embedded in Epon. Finally, ultrathin sections were stained with uranyl acetate and lead citrate, and analysed with a Tecnai Spirit electron microscope (Thermo Fisher Scientific) operated at 120 kV.

Generation of an Anti-BadA Antibody and Western Blotting

Novel rabbit anti-BadA IgG antibodies were developed using isolated BadA proteins derived from the growth culture supernatant of vortexed (2 min) *B. henselae* Marseille. BadA and other large proteins were precipitated by incubating the supernatant in 5% polyethylene glycol 6000 (Carl Roth) o/n at 4°C while slightly shaking, and subsequently collected by centrifugation at 10,000 × *g* for 1 h at 4°C. The pelleted sample was separated via sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) in a single-well 8% gel without prior heat denaturing and was stained o/n with Coomassie Blue R. BadA protein was precisely sliced from the top of the stacking gel and its identity was confirmed via mass spectrometry (see below and **Supplementary Figure 1**) and used as antigen (ca. 75 μg/injection) for generation of rabbit anti-BadA IgG antibodies (Eurogentec). Rabbit pre-immune serum was used as negative control to verify BadA antibody specificity via Western blotting. Antibodies were further purified to eliminate unspecific antibody background reaction by pre-adsorption with *B. henselae* Marseille ΔBadA-T (5 × 10^9 cells/ml) for 2 h at RT, while shaking (900 RPM).

BadA protein expression was analysed via Western blotting using the rabbit anti-BadA IgG antibodies (this study). Whole cell *B. henselae* strains were grown in BALI medium for three days, collected by centrifugation, and subsequently prepared by incubation in Laemmli sample buffer (Sigma-Aldrich) for 10 min at 95°C. Denatured proteins were separated via SDS-PAGE on a 4–15% gradient gel (Bio-Rad) and transferred to nitrocellulose membranes for 1 h at 300 mA in Towbin transfer buffer (10% glycerol). Blotted membranes were incubated o/n at 4°C with rabbit anti-BadA IgG antibodies (1:4,000) followed by a second incubation for 90 min at RT with HRP-conjugated swine anti-rabbit IgG antibodies (1:2,000; Agilent-Dako). Detected proteins were developed using SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Scientific) and analysed on a ChemiDOC XRS + system (Bio-Rad) with ImageLab V6.0.1. software (Bio-Rad).

Mass Spectrometry

The Coomassie Blue R-stained SDS-gel fragment was cut into small pieces and subsequently prepared for mass spectrometry (MS). Briefly, proteins were denatured (8 M urea and 100 mM ammonium bicarbonate) and 5 mM tris(2-carboxyethyl)phosphine hydrochloride was added to reduce the amount of disulphide bonds. 10 mM iodoacetamide was added

for alkylation in a dark room. Finally, the samples were diluted in 100 mM ammonium bicarbonate and 0.5 mg/ml sequencing-grade trypsin (Promega) was added to digest the proteins into peptides. Peptides were analysed on a Q Exactive HFX connected to an Easy-nLC 1200 instrument (Thermo Scientific).

Immunofluorescence Microscopy

Bartonella henselae BadA surface expression was assessed via immunofluorescence microscopy testing using rabbit anti-BadA IgG antibodies (this study). *B. henselae* strains grown in BALI medium were resuspended in PBS, airdried on glass microscopy slides (KNITTEL StarFrost®), and fixed with 3.75% paraformaldehyde for 10 min at 4°C. Fixed bacteria were stained with rabbit anti-BadA IgG antibodies (1:400) and subsequently incubated with goat IgG anti-rabbit IgG conjugated to Alexa 488 (1:200; Dianova), both for 1 h at RT. Bacterial DNA was stained with 4′,6-diamidino-2-phenylindole (1 μg/ml; DAPI; Merck) for 10 min at 4°C. All incubation steps were performed in a humid chamber and followed by three washes with PBS. Slides were mounted with fluorescence medium (Dako), air-dried, and analysed with a Zeiss Axio Imager 2 microscope equipped with a Spot RT3 microscope camera (Diagnostic Instruments Inc.) operated by VisiView V.2.0.5 (Visitron Systems).

Binding of *Bartonella henselae* to Extracellular Matrix Proteins

The binding of *B. henselae* to ECM proteins was tested *in vitro* using an enzyme-linked immunosorbent assays (ELISA). Briefly, 1 μg of human plasma fibronectin (Fn, Sigma-Aldrich) or human collagen-I (Col-I, Merck) was coated o/n at 4°C onto Nunc Maxisorp flat-bottom 96-wells (Thermo Scientific). Wells were blocked with 2% w/v bovine serum albumin (BSA; Sigma-Aldrich) in washing buffer (0.05% v/v Tween 20 in PBS) and incubated for 90 min at 37°C. *B. henselae* strains grown in BALI medium were resuspended in PBS, added to the wells (ca. 2×10^8 cells), and incubated for 2 h at 37°C. Bound bacteria were detected using anti-*B. henselae* IgG antibodies (1:1,000 in blocking buffer; Kempf et al., 2000) and swine anti-rabbit HRP (1:2,000 in blocking buffer). Samples were developed using 3,3′,5,5′-tetramethylbenzidine liquid substrate (TMB; Sigma-Aldrich) for ca. 1 min and the reaction was stopped with 1M HCl. The resulting absorbance was measured at 450 nm using a microplate Sunrise-Basic^{TM} reader (TECAN). All 96-well plates were sealed to prevent evaporation during incubation steps, and were each time immediately followed by three washes in wash buffer. Assays were done in triplicate.

Statistics

Statistical analyses were performed using one-way ANOVA testing on Prism V7.04 (GraphPad Software) assuming parametric data distribution. A value of $p < 0.01$ was considered statistically significant.

RESULTS

Long-Read Sequencing Demonstrates a Conserved *Bartonella henselae* Genome With Only Few Divergences in Genomic Organisation

Bartonella henselae strains and substrains from our strain collection were included in this study for which the origin was traceable. Eight complete and single contig genomes of different *B. henselae* (sub)strains (**Table 1**) were generated using next-generation long-read Pacific Biosciences SMRT sequencing. The Phred quality (Q) score of the CCS reads ranges from Q32 (99.94% accuracy) to Q34 (99.96% accuracy) for reads filtered to have a score above Q20. In case of the strains Marseille and ATCC49882^T var-1, an overall Q-score with an accuracy of 98.7% and 98.9% was observed, respectively. The genome of *B. henselae* consists of a single circular chromosome with a size that ranges from ca. 1.91 Mbp (Marseille) to 1.97 Mbp (88-64

TABLE 2 | Genome comparison as per average nucleotide identity (ANI) based on MUMmer (%).

	Marseille	ATCC49882^T var-1	ATCC49882^T var-2	Berlin-I	G-5436	88-64 Oklahoma	FR96/BK38	FR96/BK3
Marseille	–	98.83 (97.88)	98.84 (97.61)	98.84 (97.61)	98.84 (97.89)	98.83 (97.89)	98.82 (98.40)	98.91 (99.06)
ATCC49882^T var-1	98.84 (96.17)	–	99.99 (99.53)	99.98 (99.53)	99.98 (100.00)	99.92 (98.89)	99.37 (98.19)	98.60 (97.36)
ATCC49882^T var-2	98.85 (96.95)	99.99 (99.99)	–	99.99 (100.00)	99.99 (100.00)	99.92 (99.70)	99.37 (98.17)	98.61 (98.14)
Berlin-I	98.85 (96.94)	99.98 (99.99)	99.99 (100.00)	–	99.99 (100.00)	99.92 (99.70)	99.37 (98.17)	98.62 (98.14)
G-5436	98.84 (96.17)	99.99 (99.99)	99.99 (99.53)	99.99 (99.53)	–	99.92 (98.90)	99.39 (98.21)	98.61 (97.26)
88-64 Oklahoma	98.83 (96.31)	99.92 (99.96)	99.92 (99.72)	99.92 (99.72)	99.92 (99.97)	–	99.39 (97.23)	98.60 (98.38)
FR96/BK38	98.82 (96.99)	99.38 (98.45)	99.38 (97.98)	99.38 (97.98)	99.39 (98.45)	99.40 (97.45)	–	98.58 (97.46)
FR96/BK3	98.89 (97.75)	98.60 (97.74)	98.61 (97.40)	98.61 (97.40)	98.60 (97.74)	98.60 (97.75)	98.57 (97.58)	–

The upper value represents the genome comparison as per ANI (%) while the value in parentheses represents the percentage of aligned nucleotides. ANI is visualised with a colour scale gradient ranging from yellow to red (i.e., lower to higher ANI).

Oklahoma) and a low GC-content of 38%. All analysed genomes start with the housekeeping gene *gltA* (sense) and genes encoding for *Bartonella* adhesin A (BadA), the Trw locus, the VirB/D4 locus, and Pap31 (Lu et al., 2013) were identified in all strains. An overview of general sequencing parameters and genome features is given in **Supplementary Table 1**.

Whole-genome comparison as per ANI results in a score ranging from 98.57 to 99.99% indicating a low intra-species genome diversity (**Table 2**). A multiple genome alignment using progressiveMAUVE shows a conserved core genome with only few divergences in genomic organisation (**Figure 1**). Most genomic rearrangements are located in the first ca. 300,000 bp where a dispersed array of collinear and inversed regions is observed. This region corresponds to a previously described type II secretion system island including phage genes potentially linked to genomic variability (Alsmark et al., 2004; Engel and Dehio, 2009). In that same variable region, one or two (depending on the strain) incomplete but potential prophage sequences are predicted using PHASTER software (data not shown).

Further downstream from ca. nucleotide (nt) 400,000 to 1,850,000, a major genomic inversion around two adjacent collinear regions is identified in strains G-5436, 88-64 Oklahoma, and FR96/BK38 (**Figure 1**). Both sides of the inversion are characterised by a copy of the highly conserved *tuf* gene (elongation factor Tu; EF-Tu), that is flanked by either *fusA* (elongation factor G; EF-G), or genes encoding for ribosomal and transcription-related proteins possibly forming a transcriptional unit.

Based on the ANI results (**Table 2**) and the multiple genome alignment (**Figure 1**), strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, 88-64 Oklahoma, and FR96/BK38 are classified as a separate *B. henselae* subgroup. Both strains Marseille (genotype II) and FR96/BK3 appear to be genetically distinct.

The *badA* Genomic Island Consists of a Variable *badA* Gene Flanked by *badA* Pseudogenes and Conserved Up- and Downstream Regions

A noteworthy low similarity profile is observed around the location of the *badA* gene indicating a high(er) sequence variability in the otherwise rather conserved purple region (**Figure 1**). The chromosomal region containing *badA* (**Figure 2** and **Supplementary Table 3**) demonstrates a variable sequence length of *badA* (8.7 to 13.2 kb) among the various strains. In strains ATCC49882^T var-1 and Berlin-I a frameshift mutation causes a premature stop codon. The *badA* gene is preceded by a conserved *badA* pseudogene and, in case of strains ATCC49882^T var-1, G-5436, and FR96/BK38 followed by a large (ca. 21 kb) ORF disrupted by a premature stop codon with a highly similar domain organisation to *badA*. Depending on the *B. henselae* strain, four or five ORFs are identified in between the *badA* pseudogene and the *badA* gene. The ORF directly upstream of *badA* annotates as a surface protein and shares homology with TAAs, for strains Marseille and FR96/BK3, this applies to the last two ORFs.

In strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and FR96/BK38, a unique repeat sequence is observed in the *badA* island with a repeat unit length of 18 nt (5′-ART GGC GGA AGC AAY GGY-3′), and a number of repeat units of either 26 or 51. The authenticity of this sequence motif was verified by PCR (primers Repeat_Fw and Repeat_Rv) and Sanger sequencing (data not shown). Interestingly, only in strain ATCC49882^T var-2 this repeat sequence is located in the expressed BadA protein.

The *badA* island is flanked up- and downstream by a conserved gene array in all eight *B. henselae* strains. Genome annotation of this genomic island and flanking regions by both RASTtk and NCBI pipelines were individually checked *in silico* to ensure accurate and complete gene annotations.

The following genes which are located in the region ca. 25 kb upstream of the *badA* pseudogene potentially influence BadA expression: (i) an iron response regulator (*irr*; 504 bp) showing high sequence similarity to a ferric uptake repressor-like protein (Fur; for example with *irr* of *B. bacilliformis* ATCC 35685) located ca 23.5 kb upstream of the *badA* island, (ii) a resistance-nodulation-cell division (RND) efflux inner membrane transporter subunit (3,135 bp), and (iii) several ribosomal subunit proteins. In addition, ca. 10 kb upstream of the *badA* pseudogene, a conserved phosphoenolpyruvate-protein phosphotransferase gene (*ptsP*; 2,505 bp) is present. Further, a mobile genetic element (480 bp) is located ca. 3.4 kb upstream of the *badA* pseudogene and shows high sequence similarity among all strains (>99%), except for strain FR96/BK3 that displays a shorter ORF (201 bp) due to a premature stop codon. The deduced protein shows a high degree of sequence similarity with three proteins from the AAA + ATPase (ATPase associated with diverse cellular activities) superfamily that are all, among other functions, involved in DNA transposition, recombination-dependent replication, and damage repair (i.e., the MuB transposition protein, the RuvB-like protein 2, and the replication-associated recombination protein A). Following the mobile genetic element, a peculiar drop in sequence similarity is observed for strain Marseille, caused by a ca. 1.2 kb deletion. Overall, while strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma are fully identical in the region (ca. 25 kb) upstream of the *badA* island, several small genomic variations (e.g., point mutations and short deletions or insertions) are observed in the strains Marseille and FR96/BK3, indicating more frequent recombination events.

Similar to the upstream region, strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma show an identical downstream region (ca. 20 kb). Small genomic variations (e.g., point mutations and short deletions or insertions) appear only in strains Marseille, FR96/BK3 and occasionally in FR96/BK38 (not depicted). The immediate downstream ORF (573 bp) of *badA* (or the downstream *badA*-like domain region in case of strains ATCC49882^T var-1, G-5436, and FR96/BK38) shows a high sequence similarity with the invasion associated locus B (*ialB*) gene of *Bartonella* spp. A frameshift mutation in strains Marseille and FR96/BK3 shifts the transcription start site of this *ialB* homologue downstream with 16 bp. The proximity of the *ialB* homologue to *badA* suggests a single operon and subsequent protein association. In addition, the *B. henselae ialB*

gene (Deng et al., 2016) is found ca. 13.5 kb downstream of the *badA* island.

The *badA* Pseudogene Shows Low Intraspecies Variations and Has a Similar Architecture to *badA*

A *badA* pseudogene is found upstream of *badA* in all eight *B. henselae* strains (**Figure 2** and **Supplementary Table 3**) showing a similar length (5.2 kb to 5.4 kb) and a high sequence similarity (80–100%). Strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma show an identical *badA* pseudogene sequence, confirming their close relation. The putative C-terminal region of the corresponding protein (ca. 550 aa) including the putative anchor domain is identical among all eight strains analysed in this study.

The *badA* pseudogene sequence is considerably shorter than the *badA* gene but translates to a similar TAA architecture, including a head domain, a neck/stalk region, and an anchor domain (not shown). Certain regions in the *badA* pseudogenes align in sequential but interspersed order to regions in the *badA* genes with similarities ranging from 50 to 83%. The highest sequence similarities are found in the sequence for the anchor domain. Thus far, expression of the *badA* pseudogene is not analysed yet. The *badA* promoter is predicted to be located in a region ca. 350 bp upstream of the transcription start site (Okaro et al., 2020), and thus downstream of the *badA* pseudogene.

Three *Bartonella henselae* Strains Contain a Large Downstream *badA*-Like Domain Region

Three *B. henselae* strains (ATCC49882^T var-1, G-5436, and FR96/BK38) share an enormous *badA*-like domain region directly downstream of *badA* (**Figure 2** and **Supplementary Table 3**). This region encodes the typical C-terminal anchor domain sequence, as well as numerous repeat domains forming tandem arrays of repeat units, as similarly observed in *badA*. However, no distinguishable head domain is observed. A single bp mutation causes a premature stop in the otherwise perfect ORF coding for a *badA*-like domain region of either 7,128 aa (for strains ATCC49882^T var-1 and G-5436) or 6,834 aa (for strain FR96/BK38). ATCC49882^T var-1 and G-5436 share an identical *badA*-like domain region (with the exception of 2 bp), while a double deletion of ca. 600 bp in strain FR96/BK38 results in a shorter ORF and three instead of two unique 18 nt repeat sequence motifs (see above).

Bartonella Adhesin A Shows a High Intraspecies Variability Linked to Recombination Events

All eight studied *B. henselae* genomes include a *badA* gene (of which two are disrupted by an early stop codon) within their respective *badA* genomic island (**Figure 2**) and each of the deduced proteins show the characteristic TAA architecture consisting of an N-terminal head domain, a long and repetitive neck/stalk region, and a conserved C-terminal anchor domain (Szczesny and Lupas, 2008). Despite their overall similarities, a great variety in length, expression, and neck/stalk domain order is observed.

The number of distinguishable domains in the neck/stalk region ranges from 18 in strains ATCC49882^T var-2 and Berlin-I, to 34 in strain FR96/BK3, and domain sizes vary between 68 aa and 147 aa (**Figure 3A**). Among the studied strains, all BadA domains in the neck/stalk region fall into distinct categories based on pairwise similarities in protein sequence (visualised by similar colours), except for domain 14, 15, and 16 in strains ATCC49882^T var-2 and Berlin-I. This domain composition illustrates the repetitive nature of the BadA protein and the variability among the different *B. henselae* strains.

The repetitive *badA* structure complicates gene assembly based on short-read sequencing data. Using long-read sequencing, the full *badA* gene of the well-studied strain Marseille was covered in single reads, resulting in a 11,922 bp gene (**Supplementary Table 3**). The expressed protein consists of 3,973 aa (including the signal sequence of 47 aa) containing 30 neck/stalk domains (GenBank: MK993576.1). As such, two repeated motifs, each consisting of four domains, were previously overlooked (GenBank: DQ665674.1) by Sanger sequencing (Riess et al., 2004).

The neck/stalk region of *badA* among the *B. henselae* strains is highly variable compared to the more conserved head and membrane anchor sequences (**Figure 3B**). Furthermore, three types of *badA* anchor domain sequences (89 aa) were detected (**Figure 3C**). Type 1 is observed in strains ATCC49882^T var-1, G-5436, and FR96/BK38 and shows high sequence similarity with the other two anchor domain types (37 and 44 bp differences, respectively). Type 2 is present in the strains Berlin-I, ATCC49882^T var-2, and 88-64 Oklahoma and is identical to the anchor domain of the downstream *badA*-like domain region. Type 3 is found in strains Marseille and FR96/BK3 and shows 16 bp mutations (resulting in 6 aa differences) compared to type 2.

These observations suggest the occurrence of recombination events in *badA* and the *badA* island in general. For instance, strains ATCC49882^T var-1, G-5436, 88-64 Oklahoma, and FR96/BK38 show similar *badA* sequence arrangements, while strains Berlin-I and ATCC49882^T var-2 display a considerably shorter and divergent *badA* sequence, including the unique 18 nt-long repeat sequence motif within domain 15 (**Figure 3A**). The *badA* gene of the latter two strains seems the result of a fusion between the highly conserved N-terminal *badA* (up to 69 bp before the start of domain 4) and a formerly present C-terminal *badA*-like domain region (last 6,360 bp), as seen in strains G-5436 and ATCC49882^T var-1, showing an overlap of 39 bp. A similar recombination in the *badA* island in strain 88-64 Oklahoma connecting the end of *badA* domain 26 with the last 697 bp of the *badA*-like domain region may have resulted in this current *badA* gene. Moreover, it is likely that a recombination event including a former *badA*-like domain region anchor sequence and part of the N-terminal *badA* sequence previously resulted in a common *badA* gene for strains Marseille and FR96/BK3. Both strains probably underwent further gene rearrangements in their respective *badA* genes. Indications thereof can be found in

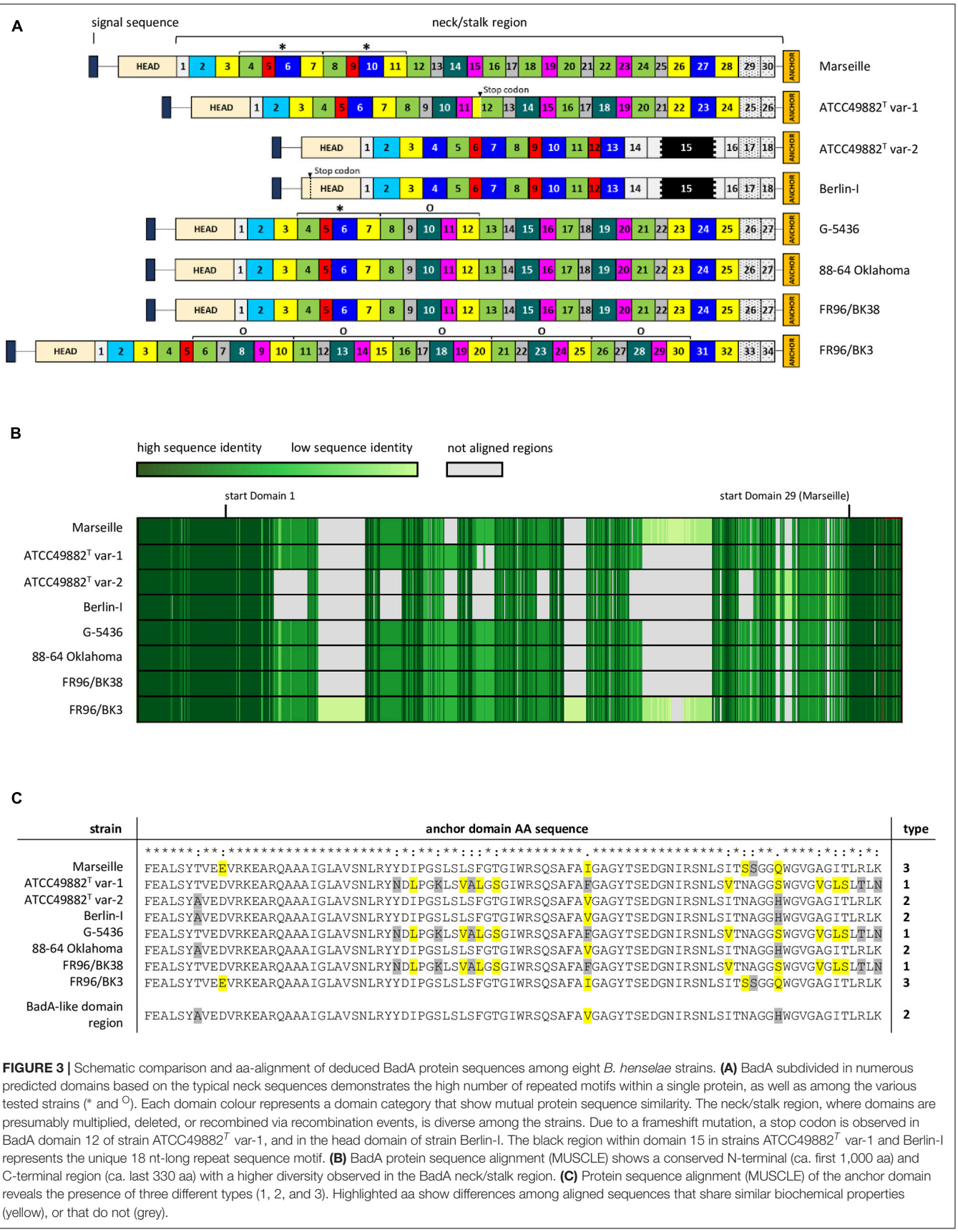


FIGURE 3 | Schematic comparison and aa-alignment of deduced BadA protein sequences among eight *B. henselae* strains. **(A)** BadA subdivided in numerous predicted domains based on the typical neck sequences demonstrates the high number of repeated motifs within a single protein, as well as among the various tested strains (* and ^{O}). Each domain colour represents a domain category that show mutual protein sequence similarity. The neck/stalk region, where domains are presumably multiplied, deleted, or recombined via recombination events, is diverse among the strains. Due to a frameshift mutation, a stop codon is observed in BadA domain 12 of strain ATCC49882^{T} var-1, and in the head domain of strain Berlin-I. The black region within domain 15 in strains ATCC49882^{T} var-1 and Berlin-I represents the unique 18 nt-long repeat sequence motif. **(B)** BadA protein sequence alignment (MUSCLE) shows a conserved N-terminal (ca. first 1,000 aa) and C-terminal region (ca. last 330 aa) with a higher diversity observed in the BadA neck/stalk region. **(C)** Protein sequence alignment (MUSCLE) of the anchor domain reveals the presence of three different types (1, 2, and 3). Highlighted aa show differences among aligned sequences that share similar biochemical properties (yellow), or that do not (grey).

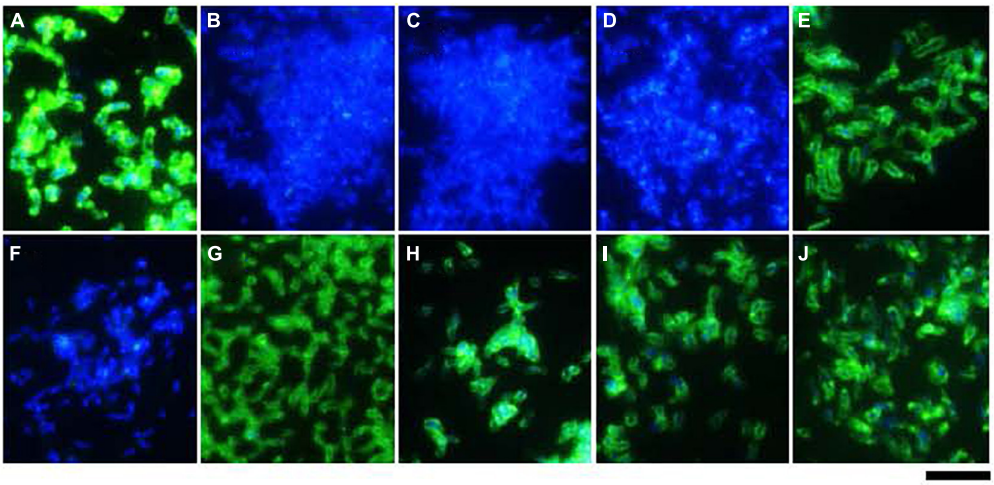


FIGURE 4 | Surface expression of BadA in various *B. henselae* strains (immunofluorescence microscopy). Surface expression of BadA was analysed via immunofluorescence microscopy using specific anti-BadA IgG antibodies (green). Bacterial DNA was counterstained using DAPI (blue). The illustrated *B. henselae* strains are **(A)** Marseille, **(B)** Marseille ΔBadA-T, **(C)** Marseille ΔBadA-D, **(D)** ATCC49882^T var-1, **(E)** ATCC49882^T var-2, **(F)** Berlin-I, **(G)** G-5436, **(H)** 88-64 Oklahoma, **(I)** FR96/BK38, and **(J)** FR96/BK3. Expression is observed for strains Marseille, ATCC49882^T var-2, G-5436, 88-64 Oklahoma, FR96/BK38, and FR96/BK3, detected by the characteristic green halo. Strains ATCC49882^T var-1 and Berlin-I do not express BadA, nor do the negative control strains Marseille ΔBadA-T and Marseille ΔBadA-D. Scale bar: 5 μm.

multiplied neck/stalk domain segments seen in BadA of strain G-5436 which are found to be repeated two and five times in BadA of strains Marseille and FR96/BK3, respectively (**Figure 3A**).

Surface Expression and Length of BadA Variants Correlate With Their Respective Genomic Sequence

Exact strain origin, passage number, BadA expression status, and BadA length are important characteristics often unknown or omitted in studies involving *B. henselae* (**Table 1**). Therefore, we performed a detailed analysis of the BadA surface expression status of eight *B. henselae* strains. Based on the long-read sequencing data, frameshift mutations in strain ATCC49882^T var-1 (caused by the deletion of a 262-bp region, at position 5,266 bp from the start codon of *badA*) and in strain Berlin-I (caused by a single bp deletion; at position 584 bp from the start codon of *badA*) both result in a premature stop codon (**Figure 2**). The *badA* genes of the other strains appeared intact.

Expression of BadA was assessed using immunofluorescence (**Figure 4**) and transmission electron microscopy (TEM; **Figure 5**). Transposon mutants Marseille ΔBadA-T and Marseille ΔBadA-D were used as negative controls. Surface-expressed BadA was clearly shown for strains Marseille, ATCC49882^T var-2, G-5436, 88-64 Oklahoma, FR96/BK38, and FR96/BK3. Strains ATCC49882^T var-1 and Berlin-I do not express BadA, similar to both negative control strains. These data were corroborated by TEM where two different preparation methods (PLT in DMF and HPF/FS) were used to achieve optimal visualisation of BadA.

The average length of BadA of the particular *B. henselae* strains was analysed from transmission electron microscopy (TEM) images (**Figure 6A**). The lengths of the BadA fibres are in accordance with their identified *badA* gene lengths. Strains Marseille (BadA; 3,973 aa) and FR96/BK3 (BadA; 4,407 aa) express the longest fibres measuring 243 and 238 nm on average, respectively, while strain ATCC49882^T var-2 (BadA; 2,920 aa) presents the shortest fibres with an average length of 155 nm. Strains G-5436 (BadA; 3,641 aa), 88-64 Oklahoma (BadA; 3,643 aa), and FR96/BK38 (BadA; 3,641 aa) display comparable fibres with average lengths of 171, 186, and 166 nm, respectively.

Finally, the molecular weights (MW) and expression of the various BadA proteins were visualised via Western blotting using rabbit anti-BadA IgG antibodies (**Figure 6B**). Blotting results are in line with immunofluorescence and electron microscopy results, as well as with MW predictions of the monomeric BadA proteins (ranging from 296 to 464 kDa). Although the resolution of SDS-PAGE does not allow for a precise MW quantification of the trimeric BadA protein, strains Marseille and FR96/BK3 show a slightly larger monomeric BadA protein (417 and 464 kDa, respectively) compared to the other strains. The uppermost detected protein is presumed to be trimeric BadA that is unable to travel down in the gel because of its enormous size, heat stability, and incomplete denaturation (Grosskinsky et al., 2007).

The Surface Expression of BadA Determines Extracellular Matrix Binding Independent of Strain Specific Domain Architecture

The biological function of the varying BadA proteins in binding of *B. henselae* to various matrix proteins was assessed. For this purpose, bacterial adhesion to human plasma Fn and human Col-I was quantified via a whole-cell ELISA using anti-*B. henselae* IgG antibodies. Surface-expressed BadA proteins

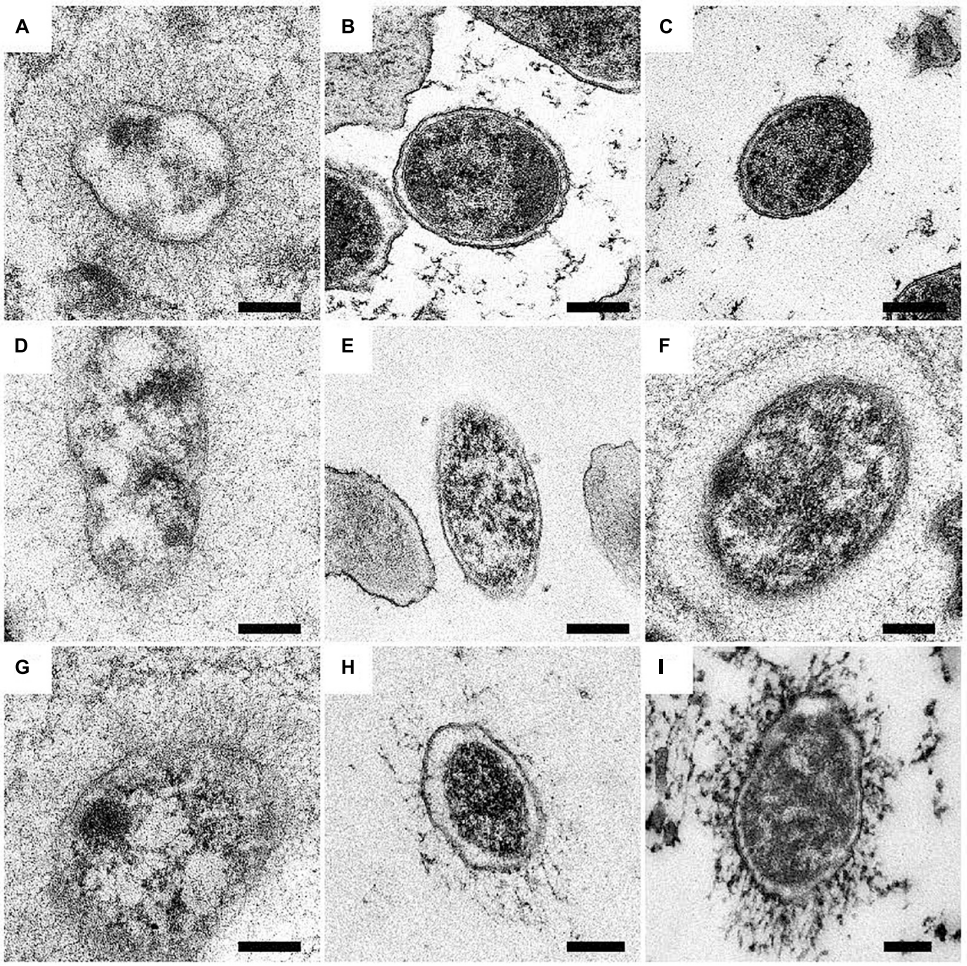


FIGURE 5 | Surface expression of BadA in various *B. henselae* strains (transmission electron microscopy). Representative images of *B. henselae* **(A)** Marseille, **(B)** Marseille ΔBadA-T, **(C)** ATCC49882^T var-1, **(D)** ATCC49882^T var-2, **(E)** Berlin-I, **(F)** G-5436, **(G)** 88-64 Oklahoma, **(H)** FR96/BK38, and **(I)** FR96/BK3 are visualised with TEM. Around 30-50 TEM images of bacterial cells per strain were analysed for determination of the phenotype. Expression of BadA is observed for strains Marseille, ATCC49882^T var-2, G-5436, 88-64 Oklahoma, FR96/BK38, and FR96/BK3 but not for strains ATCC49882^T var-1 and Berlin-I, nor for the negative control strain Marseille ΔBadA-T. For technical reasons, samples were prepared by both PLT in DMF and K4M embedding (e.g., Marseille, ATCC49882^T var-2, G-5436, and 88-64 Oklahoma), and HPF/FS and Epon embedding (e.g., Marseille ΔBadA-T, ATCC49882^T var-1, Berlin-I, FR96/BK38, and FR96/BK3). Scale bar: 200 nm.

clearly mediate bacterial binding to human plasma Fn and human Col-I (**Figure 7**). A significantly lower adhesion was observed for the BadA-deficient strains ATCC49882^T var-1, Berlin-I, and both negative control strains Marseille ΔBadA-T and Marseille ΔBadA-D. In contrast, ECM adhesion was detected for all strains that express BadA naturally, independent of their individual BadA domain composition, repeats, or various BadA arrangements.

DISCUSSION

Adhesion to cells is the first and most decisive step in the course of any infection. TAAs represent a major class of pathogenicity factors in Gram-negative bacteria and are defined by their homologous C-terminal membrane anchor (Linke et al., 2006). The best examined TAA is the *Yersinia* adhesin A (YadA) of *Yersinia enterocolitica*, and despite its relatively short gene sequence (1.3 to 1.5 kb) and small fibre length (ca. 23 nm), it is seen as the prototypical TAA (El Tahir and Skurnik, 2001). Examples of other well-studied TAAs are the *Acinetobacter* trimeric autotransporter (Ata) of *A. baumannii* (Weidensdorfer et al., 2019) and the *Neisseria* adhesin A (NadA) of *N. meningitidis* (Comanducci et al., 2002). TAAs share a modular and repetitive domain composition (head, neck, stalk, and membrane anchor domains) (Szczesny and Lupas, 2008) and the underlying repeats on the DNA level suggest easy gene rearrangements via recombination, presumably driven by genetic adaptation to alternating host environments (Lindroos et al., 2006). TAAs of the genus *Bartonella* stand out because of their variability in length, high conservation within the genus, and the presence of multiple gene variants within one genome. The

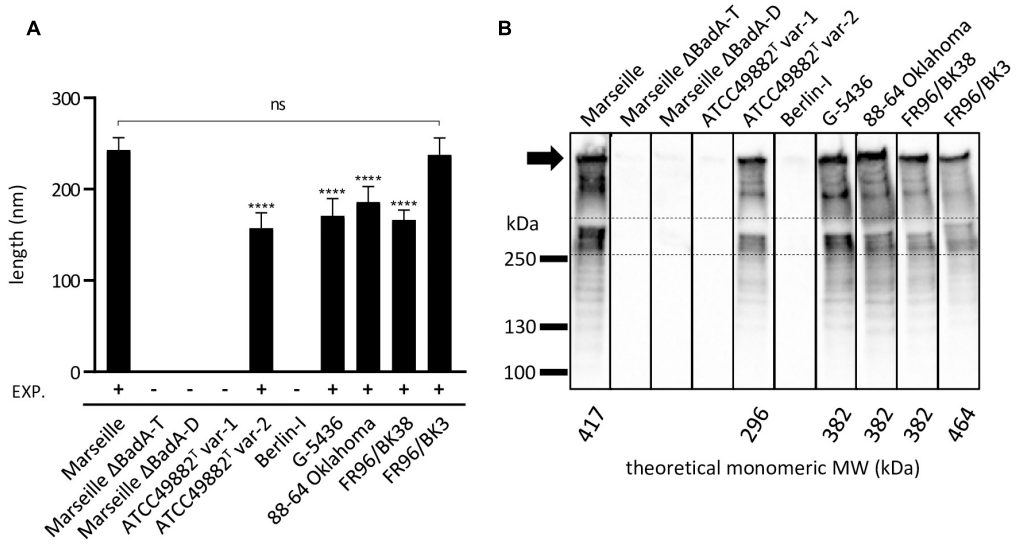


FIGURE 6 | Length analysis of surface-expressed BadA and their corresponding molecular weight of *B. henselae* strains. **(A)** The average length of expressed BadA was measured using 30-50 TEM images of bacterial cells per strain. The individual average BadA fibre lengths of strains ATCC49882^T var-2 (157 nm), G-5436 (171 nm), 88-66 Oklahoma (186 nm) and FR96/BK38 (166 nm) are significantly shorter (****$p < 0.0001$) compared to those of strains Marseille (243 nm) and FR96/BK3 (238 nm). The length of the last two strains do not differ significantly (ns). Strains ATCC49882^T var-1, Berlin-I, Marseille ΔBadA-T and Marseille ΔBadA-D do not express BadA. EXP.: expression of BadA. **(B)** BadA expression of the various *B. henselae* strains is visualised by Western blotting. The band between the dashed lines is considered monomeric BadA protein. Marseille and FR96/BK3 display a higher MW band around the predicted monomeric BadA protein. The numerous lower MW-bands are considered degradation products of the high MW BadA protein. The uppermost band is presumably trimeric BadA protein stuck in the pocket of the gel (black arrow). Bands in between the trimeric and monomeric protein are also considered degradation products of trimeric BadA. All strains were analysed on a single nitrocellulose blot, the order of columns has been rearranged *in silico*.

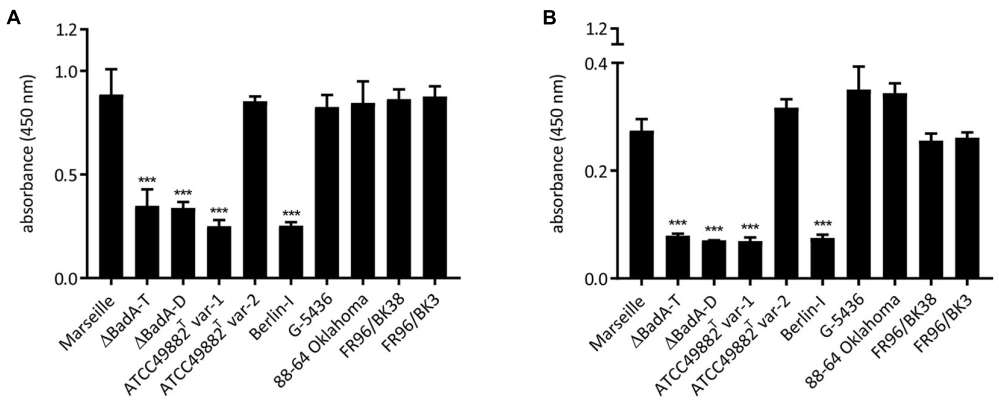


FIGURE 7 | Adhesion of *B. henselae* strains to extracellular matrix proteins. Binding of the various *B. henselae* strains to **(A)** human plasma Fn and **(B)** human Col-I. Wells were coated with Fn and Col-I and adherent bacteria were quantified via whole cell ELISA (see section "Materials and Methods"). BadA-expressing strains show a significantly higher binding to human plasma Fn and human Col-I compared to strains lacking BadA expression. Statistical significance was determined using one-way ANOVA testing (***$p < 0.001$).

extensive number (30) of repetitive neck/stalk domains in BadA of *B. henselae* Marseille makes this region an excellent "toolbox" for TAA evolution (**Figure 3A**).

The presence of numerous prophages and genomic islands including both short tandem repeats and longer repeats (such as the *badA* island of *B. henselae*) often result in sequencing and assembly errors. Long-read WGS techniques overcome these issues (Koren et al., 2013; Tørresen et al., 2019). Accordingly, we demonstrate that the length of *badA* from strain Marseille is 11,922 bp (**Supplementary Table 3**), instead of the earlier determined 9,249 bp (Riess et al., 2004). In addition, long-read sequencing of the *badA* gene from seven other *B. henselae* strains confirmed the previously described varying length of the neck/stalk region among the various *B. henselae* strains (**Figure 3B**; Riess et al., 2007).

All sequenced *B. henselae* strains show a conserved genome with a high intra-species genome sequence similarity. Divergences are mostly observed in a 300 kb region that includes a defined type II secretion system island and contains various phage genes (Engel and Dehio, 2009). Within this region which contains highly dynamic stretches, genomic inversions were identified (e.g., around the duplicated *tuf* gene and multiple copies of rRNA operons). This prophage region presumably drives diversification and dispersion of specific host-adaptability genes throughout the *B. henselae* population (Anderson et al., 1994; Lindroos et al., 2006; Harvey et al., 2019). Such diverse recombination events make correct species typing or phylogenetic analysis challenging.

Two *B. henselae* genotypes have been described so far. Genotype I is believed to be more associated with human infections, while genotype II appears to outcompete genotype I during bloodstream infections in cats (Huwyler et al., 2017). Strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma are all human isolates with high pairwise genome sequence identity of >99.9% (**Table 2**). In addition, three of these strains are known variants of the type-strain ATCC49882^T Houston-I (**Table 1**). Based on our data, we suggest to classify these five strains as genotype I. Strain Marseille shows a comparatively lower genome sequence identity (ca. 98.8%) to these five strains (**Table 2**) and demonstrates that genotype II strains are not exclusively isolated from cats. Compared to the genotype I group, both *B. henselae* cat isolates in this study (strains FR96/BK38 and FR96/BK3) do not show a high pairwise genome sequence identity (98.6%). We propose that the classification of *B. henselae* strains into particular genotypes cannot be solely attributed to the source of isolation. This is in accordance with findings that new genetic variants of *B. henselae* frequently emerge *in vivo* (Iredell et al., 2003; Berghoff et al., 2007).

Based on the various hypothesised rearrangements within the defined *badA* island sequence among the different *B. henselae* strains, including the downstream *badA*-like domain region and the small variations in the BadA anchor domain aa-sequence, we postulate that a series of recombination events has led to the formation of at least three phylogenetically distinct groups of *B. henselae* (**Figure 3C**). We propose strain G-5436 to be the evolutionary ancestor of the strains in this study because of its intact *badA* sequence and its present (and longer) downstream *badA*-like domain region (compared to strain FR96/BK38). Future classification and bacterial genotyping of *B. henselae* strains should be supported by either long-read WGS results or a detailed analysis of the *badA* island.

As known so far, *badA* gene expression seems to be regulated by (i) the general stress response system (Tu et al., 2016), (ii) the *Bartonella* regulatory transcript in combination with the transcriptional regulator protein (Tu et al., 2017), and (iii) the BatR/S two-component system (Quebatte et al., 2010). These systems are influenced by the extracellular environment, correlating with frequently alternating host conditions (e.g., temperature and pH). The close vicinity of a predicted ferric uptake repressor-like protein (Fur) to *badA* suggests a repressed *badA* expression of *B. henselae* while residing in the iron-saturated flea gut environment. In mammal hosts where free heme as an iron source is rare, upregulation of *badA* might facilitate initial adhesion and subsequent infection (Battisti et al., 2007). An upstream gene (*ptsP*) that is part of a phosphotransferase (PT) system, controls among others the carbohydrate flow in bacteria and is involved in virulence gene expression related to nutrient availability (Bier et al., 2020). An *ialB* homologue of *B. henselae* is located immediately downstream of the *badA* genomic island. It has been described that IalB plays a role in erythrocyte invasion of *Bartonella* species (Coleman and Minnick, 2001; Deng et al., 2016). However, and in contrast to *badA*, *ialB* expression is upregulated by a low pH or decreasing temperatures (Coleman and Minnick, 2003; Okaro et al., 2017). A similar operon configuration involving a TAA gene associated with *ialB* is observed for the *Brucella abortus* trimeric autotransporter adhesin (BatA; Rahbar et al., 2020). The functional relationship of *badA* with Fur, the PT system, and the IalB homologue remains to be explored.

Microscopy and protein expression analyses validate the long-read sequencing results and confirm that all strains express a surface-expressed BadA, except for strains ATCC49882^T var-1 and Berlin-I due to individual frameshift mutations (**Figures 4–6**). The length of surface-expressed BadA was determined from TEM-images and correlates with the *badA* gene length of the respective *B. henselae* strain. The slight variation between the measured fibre lengths of surface-expressed BadA from *B. henselae* strains G-5436, 88-64 Oklahoma, and FR96/BK38, despite having similar *badA* gene lengths, might derive from structural changes due to the TEM-processing, such as partial protein unfolding (**Figure 6** and **Supplementary Table 3**).

BadA expression was previously shown to be crucial for endothelial cell infection and induction of a proangiogenic response (Riess et al., 2004) while negatively affecting VirB/D4-dependent host cell invasion via "invasome" formation (Lu et al., 2013). Binding of *B. henselae* to ECM proteins human plasma Fn and Col-I also depends on the expression of BadA rather than the observed differences in protein length (**Figure 7**). The binding of BadA to Fn has been previously attributed to the passenger domain, while the head domain itself was described to be sufficient for collagen binding. In addition, it was demonstrated that a critical length of surface-expressed BadA (40 nm) was necessary to detect binding of *B. henselae* to Fn (Kaiser et al., 2008, 2012). Plasma Fn which is present in blood and other fluids, is a major component of the fibrin clot in early wound healing responses and an important glycoprotein in the ECM. Col-I is the most abundant collagen in the human body, especially in the dermis, and represents a major binding partner for TAAs (Łyskowski et al., 2011; Vaca et al., 2019). In case of a cat scratch or bite in the human dermis, both ECM proteins might thus be an ideal BadA binding partner in the course of infection of *B. henselae*.

The frequent transitions of *B. henselae* from the cat flea's gut to cats and accidental human hosts might require efficient and quick adaptation strategies. In contrast to other variable regions, the *badA* island has previously not been predicted nor described as a prophage or defined genomic island

(Engel and Dehio, 2009). The observed heterogeneity in the *badA* island between different *B. henselae* strains might be the result of evolutionary selection for various beneficial traits such as host-specific colonisation and immune escape. We postulate a shuffling mechanism comparable to phase variation that capitalises on the extensive number of repeats in the *badA* island making this region an excellent "toolbox" for TAA evolution. Combinatorial reshuffling of similar TAA repeats as a mechanism to quickly adapt these immunodominant adhesins has been proposed previously (Szczesny et al., 2008). The highly repetitive *badA* neck/stalk region might facilitate site-specific recombination or slipped-strand mispairing within the *badA* island that reshuffles *badA*, the neighbouring *badA* pseudogenes, and possibly involves the peculiar 18 nt-long repeat sequence motif. The protein structure of this translated 18 nt-long repeat sequence might mimic collagen-like triple helixes which would fit the general TAA structure. Similar phase variation mechanisms were demonstrated earlier for *N. meningitidis* (Bentley et al., 2007). A potential candidate that mediates these recombination events is the identified mobile genetic element protein upstream of the *badA* island. This ORF shows homology with three different AAA + ATPases (i.e., MuB, RuvB-like protein 2, and RarA) that are all involved in DNA recombination. The presence of *B. henselae* isolates lacking BadA expression due to a frameshift mutation (e.g., strain ATCC49882^T var-2 and Berlin-I) or the loss of BadA expression following multiple passages outside a natural host, caused by the deletion of an 8.5 kb genomic fragment (Kempf et al., 2001; Riess et al., 2004), strengthens our hypothesis of existing phase-on and phase-off phenotypes (Anderson and Neuman, 1997; Kyme et al., 2003; Lu et al., 2013). Because of the unknown passage number of the respective bacterial strains (before arriving to our laboratory) we cannot fully rule out that such recombination events might have occurred while cultivating the bacteria under laboratory conditions. However, regaining the ability to express BadA was, to our knowledge, never described. Irreversible loss of biologically important genetic information might also occur *in vivo* upon changing environments. Such presumably attenuated or non-pathogenic mutants might stay restricted, e.g., to the vector niche because they lost the ability to infect the host. They might even not survive in nature but only under artificial laboratory conditions where the metabolic burden of BadA expression would select for faster growing bacteria that silence the *badA* gene (e.g., via deletion or single bp mutations). However, this hypothesis has not been proven for *B. henselae* because suitable animal infection models mimicking human infections do not exist.

Overall, *badA* domain duplications, acquisitions, and deletions result in varying BadA proteins with different lengths and alternating repetitive domain arrangements, as is demonstrated within this rather small data set of eight *B. henselae* isolates. The diversity of BadA proteins illustrates the occurrence of active recombination events within the *badA* island, demonstrates the utilisation of the downstream *badA*-like domain region acting as a "toolbox" for rearrangements in *badA* composition, and strengthens our hypothesis of active phase variation via site-specific recombination in the scope of rapid host adaptation (e.g., pathogenicity and immune evasion).

CONCLUSION

A detailed analysis of long-read sequenced genomes is required to discern subtle differences and numerous repeat sequences between closely related bacterial genotypes. *B. henselae* and BadA provide an excellent setting to analyse host adaptation and evolution of pathogenic proteins and might serve as an example for other TAAs. BadA is verified to be important for the initial host-pathogen interaction via binding to ECM proteins (e.g., Fn and Col-I). It is thus important to verify surface expression of BadA (or other TAAs of *Bartonella* spp.) before starting infection experiments. By studying the diversity of eight *B. henselae* strains (human and cat isolates), we have gathered evidence of active recombination within the repetitive *badA* genomic island and suggest that among the studied strains, *B. henselae* G-5436 is the evolutionary ancestor. The pathogenic life cycle of *B. henselae* might promote frequent recombination events leading to the combinatorial reshuffling of similar domain arrays and the emergence of variations among different BadA proteins that possibly contribute to host immune evasion and enhance long term and efficient survival in the differing host conditions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors upon request, without undue reservation. The genome sequences of all sequenced *B. henselae* isolates, together with their corresponding SRA data, have been deposited in the NCBI GenBank database under BioProject PRJNA720375 with the following genome accession numbers: CP072904 (Marseille), CP072903 (ATCC49882T var-1), CP072902 (ATCC49882T var-2), CP072901 (Berlin-I), CP072900 (G-5436), CP072899 (88-64 Oklahoma), CP072898 (FR96/BK38), and CP072897 (FR96/BK3) (**Supplementary Table 1**). The genomic sequence of badA from *B. henselae* Marseille was deposited separately under the GenBank accession number MK993576.1.

AUTHOR CONTRIBUTIONS

VK, AT, and DL designed the study. AT and WB performed the experimental work. KH provided the electron microscopy. SC and JM performed the mass spectrometry. DV contributed to binding experiments. AS helped in cloning experiments and long-read sequencing. AT and VK wrote the manuscript. All authors approved the final manuscript.

FUNDING

This research was supported by the Viral and Bacterial Adhesin Network Training (ViBrANT) Programme funded by the European Union's HORIZON 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No 765042 (to VK, JM, and DL), by the Robert Koch-Institute, Berlin, Germany (*Bartonella*

consiliary laboratory, 1369-354, to VK), and by the Deutsche Forschungsgemeinschaft (DFG FOR 2251 "Adaptation and persistence of *Acinetobacter baumannii*" to VK).

ACKNOWLEDGMENTS

The PacBio sequencing service was provided by the Norwegian Sequencing Centre (sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the Functional Genomics and Infrastructure programmes of the Research Council of Norway and the South-eastern Regional Health Authorities. Special thanks goes to Ave Tooming-Klunderud (University of Oslo, Norway) for technical help. We also thank Iris Koch (MPI Tübingen, Germany) for the EM technical processing, Andrei N. Lupas (MPI Tübingen, Germany) for the interpretation on the structure of the unique repeat sequence, and Kjetill Sigurd Jakobsen (University of Oslo, Norway) for the supportive discussion on genomic rearrangement mechanisms.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.838267/full#supplementary-material

REFERENCES

Alsmark, C. M., Frank, A. C., Karlberg, E. O., Legault, B. A., Ardell, D. H., Canbäck, B., et al. (2004). The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc. Natl. Acad. Sci. U S A.* 101, 9716–9721. doi: 10.1073/pnas.0305659101

Anderson, B. E., and Neuman, M. A. (1997). *Bartonella* spp. as emerging human pathogens. *Clin. Microbiol. Rev.* 10, 203–219. doi: 10.1128/cmr.10.2.203-219.1997

Anderson, B., Goldsmith, C., Johnson, A., Padmalayam, I., and Baumstark, B. (1994). Bacteriophage-like particle of *Rochalimaea henselae*. *Mol. Microbiol.* 13, 67–73. doi: 10.1111/j.1365-2958.1994.tb00402.x

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Arvand, M., Wendt, C., Regnath, T., Ullrich, R., and Hahn, H. (1998). Characterization of *Bartonella henselae* isolated from bacillary angiomatosis lesions in a human immunodeficiency virus–infected patient in Germany. *Clin. Infect. Dis.* 26, 1296–1299. doi: 10.1086/516348

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

Bassler, J., Hernandez Alvarez, B., Hartmann, M. D., and Lupas, A. N. (2015). A domain dictionary of *trimeric autotransporter adhesin*s. *Int. J. Med. Microbiol.* 305, 265–275. doi: 10.1016/j.ijmm.2014.12.010

Battisti, J. M., Smitherman, L. S., Sappington, K. N., Parrow, N. L., Raghavan, R., and Minnick, M. F. (2007). Transcriptional regulation of the heme binding protein gene family of *Bartonella quintana* is accomplished by a novel promoter element and iron response regulator. *Infect. Immun.* 75, 4373–4385. doi: 10.1128/IAI.00497-07

Bayer, M. E., Carlemalm, E., and Kellenberger, E. (1985). Capsule of *Escherichia coli* K29: ultrastructural preservation and immunoelectron microscopy. *J. Bacteriol.* 162, 985–991. doi: 10.1128/jb.162.3.985-991.1985

Bentley, S. D., Vernikos, G. S., Snyder, L. A. S., Churcher, C., Arrowsmith, C., Chillingworth, T., et al. (2007). Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet.* 3:0230–0240. doi: 10.1371/journal.pgen.0030023

Berghoff, J., Viezens, J., Guptill, L., Fabbi, M., and Arvand, M. (2007). *Bartonella henselae* exists as a mosaic of different genetic variants in the infected host. *Microbiology* 153, 2045–2051. doi: 10.1099/mic.0.2007/006379-0

Berrich, M., Kieda, C., Grillon, C., Monteil, M., Lamerant, N., Gavard, J., et al. (2011). Differential effects of *Bartonella henselae* on human and feline macro- and micro-vascular endothelial cells. *PLoS One* 6:e20204. doi: 10.1371/journal.pone.0020204

Bier, N., Hammerstrom, T. G., and Koehler, T. M. (2020). Influence of the phosphoenolpyruvate:carbohydrate phosphotransferase system on toxin gene expression and virulence in *Bacillus anthracis*. *Mol. Microbiol.* 113, 237–252. doi: 10.1111/mmi.14413

Breitschwerdt, E. B. (2017). Bartonellosis, one health and all creatures great and small. *Vet. Dermatol.* 28, 96–e21. doi: 10.1111/vde.12413

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5:8365. doi: 10.1038/srep08365

Buffet, J. P., Pisanu, B., Brisse, S., Roussel, S., Félix, B., Halos, L., et al. (2013). Deciphering *Bartonella* diversity, recombination, and host specificity in a rodent community. *PLoS One* 8:0068956. doi: 10.1371/journal.pone.0068956

Chang, C.-C., Chen, Y.-J., Tseng, C.-S., Lai, W.-L., Hsu, K.-Y., Chang, C.-L., et al. (2011). A comparative study of the interaction of *Bartonella henselae* strains with human endothelial cells. *Vet. Microbiol.* 149, 147–156. doi: 10.1016/j.vetmic.2010.09.033

Chomel, B. B., Kasten, R. W., Floyd-Hawkins, K., Chi, B., Yamamoto, K., Roberts-Wilson, J., et al. (1996). Experimental transmission of *Bartonella henselae* by the cat flea. *J. Clin. Microbiol.* 34, 1952–1956. doi: 10.1128/jcm.34.8.1952-1956.1996

Coleman, S. A., and Minnick, M. F. (2001). Establishing a direct role for the *Bartonella bacilliformis* invasion-associated locus B (IalB) protein in human erythrocyte parasitism. *Infect. Immun.* 69, 4373–4381. doi: 10.1128/IAI.69.7.4373-4381.2001

Coleman, S. A., and Minnick, M. F. (2003). Differential expression of the invasion-associated locus B (ialB) gene of *Bartonella bacilliformis* in response to environmental cues. *Microb. Pathog.* 34, 179–186. doi: 10.1016/S0882-4010(03)00005-6

Comanducci, M., Bambini, S., Brunelli, B., Adu-Bobie, J., Aricò, B., Capecchi, B., et al. (2002). NadA, a novel vaccine candidate of *Neisseria meningitidis*. *J. Exp. Med.* 195, 1445–1454. doi: 10.1084/jem.20020407

Darling, A. C. E., 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

Darling, A. E., Mau, B., and Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:0011147. doi: 10.1371/journal.pone.0011147

Deng, H., Pang, Q., Xia, H., Le Rhun, D., Le Naour, E., Yang, C., et al. (2016). Identification and functional analysis of invasion associated locus B (IalB) in *Bartonella* species. *Microb. Pathog.* 98, 171–177. doi: 10.1016/j.micpath.2016.05.007

Drancourt, M., Birtles, R., Raoult, D., Chaumentin, G., Vandenesch, F., and Etienne, J. (1996). New serotype of *Bartonella henselae* in endocarditis and cat-scratch disease. *Lancet* 347, 441–443. doi: 10.1016/S0140-6736(96)90012-4

El Tahir, Y., and Skurnik, M. (2001). YadA, the multifaceted *Yersinia* adhesin. *Int. J. Med. Microbiol.* 291, 209–218. doi: 10.1078/1438-4221-00119

Engel, P., and Dehio, C. (2009). Genomics of host-restricted pathogens of the genus *Bartonella*. *Microb. Pathogenom.* 2009, 158–169. doi: 10.1159/000235769

Engel, P., Salzburger, W., Liesch, M., Chang, C.-C., Maruyama, S., Lanz, C., et al. (2011). Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen *Bartonella*. *PLoS Genet.* 7:e1001296. doi: 10.1371/journal.pgen.1001296

Ettema, T. J. G., and Andersson, S. G. E. (2009). The α-*proteobacteria*: the Darwin finches of the bacterial world. *Biol. Lett.* 5, 429–432. doi: 10.1098/rsbl.2008.0793

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 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

Grosskinsky, U., Schütz, M., Fritz, M., Schmid, Y., Lamparter, M. C., Szczesny, P., et al. (2007). A conserved glycine residue of trimeric autotransporter domains plays a key role in *Yersinia* adhesin A autotransport. *J. Bacteriol.* 189, 9011–9019. doi: 10.1128/JB.00985-07

Harms, A., and Dehio, C. (2012). Intruders below the radar: molecular pathogenesis of *Bartonella* spp. *Clin. Microbiol. Rev.* 25, 42–78. doi: 10.1128/CMR.05009-11

Harvey, K. L., Jarocki, V. M., Charles, I. G., and Djordjevic, S. P. (2019). The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. *Front. Microbiol.* 10:2351. doi: 10.3389/fmicb.2019.02351

Hildebrand, A., Remmert, M., Biegert, A., and Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins Struct. Funct. Bioinforma.* 77, 128–132. doi: 10.1002/prot.22499

Hoiczyk, E. (2000). Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J.* 19, 5989–5999. doi: 10.1093/emboj/19.22.5989

Huwyler, C., Heiniger, N., Chomel, B. B., Kim, M., Kasten, R. W., and Koehler, J. E. (2017). Dynamics of co-Infection with *Bartonella henselae* genotypes I and II in naturally infected cats: Implications for feline vaccine development. *Microb. Ecol.* 74, 474–484. doi: 10.1007/s00248-017-0936-8

Iredell, J., Blanckenberg, D., Arvand, M., Grauling, S., Feil, E. J., and Birtles, R. J. (2003). Characterization of the natural population of *Bartonella henselae* by multilocus sequence typing. *J. Clin. Microbiol.* 41, 5071–5079. doi: 10.1128/JCM.41.11.5071-5079.2003

Kaiser, P. O., Linke, D., Schwarz, H., Leo, J. C., and Kempf, V. A. J. (2012). Analysis of the BadA stalk from *Bartonella henselae* reveals domain-specific and domain-overlapping functions in the host cell infection process. *Cell. Microbiol.* 14, 198–209. doi: 10.1111/j.1462-5822.2011.01711.x

Kaiser, P. O., Riess, T., Wagner, C. L., Linke, D., Lupas, A. N., Schwarz, H., et al. (2008). The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. *Cell. Microbiol.* 10, 2223–2234. doi: 10.1111/j.1462-5822.2008.01201.x

Kempf, V. A. J., Schaller, M., Behrendt, S., Volkmann, B., Aepfelbacher, M., Cakman, I., et al. (2000). Interaction of *Bartonella henselae* with endothelial cells results in rapid bacterial rRNA synthesis and replication. *Cell. Microbiol.* 2, 431–441. doi: 10.1046/j.1462-5822.2000.00072.x

Kempf, V. A. J., Volkmann, B., Schaller, M., Sander, C. A., Alitalo, K., Rieß, T., et al. (2001). Evidence of a leading role for VEGF in *Bartonella henselae*-induced endothelial cell proliferations. *Cell. Microbiol.* 3, 623–632. doi: 10.1046/j.1462-5822.2001.00144.x

Kešnerová, L., Moritz, R., and Engel, P. (2016). Bartonella apis sp. nov., a honey bee gut symbiont of the class Alpha*proteobacteria*. *Int. J. Syst. Evol. Microbiol.* 66, 414–421. doi: 10.1099/ijsem.0.000736

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0

Koren, S., Harhay, G. P., Smith, T. P. L., Bono, J. L., Harhay, D. M., Mcvey, S. D., et al. (2013). Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* 14:r101. doi: 10.1186/gb-2013-14-9-r101

Kyme, P., Dillon, B., and Iredell, J. (2003). Phase variation in *Bartonella henselae*. *Microbiology* 149, 621–629. doi: 10.1099/mic.0.26014-0

Lamps, L. W., and Scott, M. A. (2004). Cat-scratch disease: historic, clinical, and pathologic perspectives. *Am. J. Clin. Pathol.* 121(Suppl.), 71–80. doi: 10.1309/JC8YM53L4E0L6PT5

Leo, J. C., Grin, I., and Linke, D. (2012). Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 1088–1101. doi: 10.1098/RSTB.2011.0208

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Lindroos, H., Vinnere, O., Mira, A., Repsilber, D., Näslund, K., and Andersson, S. G. E. (2006). Genome rearrangements, deletions, and amplifications in the natural population of *Bartonella henselae*. *J. Bacteriol.* 188, 7426–7439. doi: 10.1128/JB.00472-06

Linke, D., Riess, T., Autenrieth, I. B., Lupas, A., and Kempf, V. A. J. (2006). *Trimeric autotransporter adhesins*: variable structure, common function. *Trends Microbiol.* 14, 264–270. doi: 10.1016/j.tim.2006.04.005

Lu, Y. Y., Franz, B., Truttmann, M. C., Riess, T., Gay-Fraret, J., Faustmann, M., et al. (2013). *Bartonella henselae trimeric autotransporter adhesin* BadA expression interferes with effector translocation by the VirB/D4 type IV secretion system. *Cell. Microbiol.* 15, 759–778. doi: 10.1111/cmi.12070

Łyskowski, A., Leo, J. C., and Goldman, A. (2011). "Structure and biology of *trimeric autotransporter adhesins*," in *Bacterial Adhesion. Advances in Experimental Medicine and Biology*, Vol. 715, eds D. Linke and A. Goldman (Dordrecht: Springer), doi: 10.1007/978-94-007-0940-9_9

Müller, N. F., Kaiser, P. O., Linke, D., Schwarz, H., Riess, T., Schäfer, A., et al. (2011). *Trimeric autotransporter adhesin*-dependent adherence of *Bartonella henselae*, *Bartonella quintana*, and *Yersinia enterocolitica* to matrix components and endothelial cells under static and dynamic flow conditions. *Infect. Immun.* 79, 2544L–2553. doi: 10.1128/IAI.01309-10

Okaro, U., Addisu, A., Casanas, B., and Anderson, B. (2017). *Bartonella* species, an emerging cause of blood-culture-negative endocarditis. *Clin. Microbiol. Rev.* 30, 709–746. doi: 10.1128/CMR.00013-17

Okaro, U., George, S., Valdes, S., Macaluso, K., and Anderson, B. (2020). A non-coding RNA controls transcription of a gene encoding a DNA binding protein that modulates biofilm development in *Bartonella henselae*. *Microb. Pathog.* 2020:104272. doi: 10.1016/j.micpath.2020.104272

Quebatte, M., Dehio, M., Tropel, D., Basler, A., Toller, I., Raddatz, G., et al. (2010). The BatR/BatS two-component regulatory system controls the adaptive response of *Bartonella henselae* during human endothelial cell infection. *J. Bacteriol.* 192, 3352–3367. doi: 10.1128/JB.01676-09

Rahbar, M. R., Zarei, M., Jahangiri, A., Khalili, S., Nezafat, N., Negahdaripour, M., et al. (2020). Non-adaptive evolution of trimeric autotransporters in *Brucellaceae*. *Front. Microbiol.* 11:2664. doi: 10.3389/fmicb.2020.560667

Regnery, R. L., Anderson, B. E., Clarridge, J. E., Rodriguez-Barradas, M. C., Jones, D. C., and Carr, J. H. (1992). Characterization of a novel *Rochalimaea* species, *R. henselae* sp. nov., isolated from blood of a febrile, human immunodeficiency virus-positive patient. *J. Clin. Microbiol.* 30, 265–274. doi: 10.1128/jcm.30.2.265-274.1992

Relman, D. A., Loutit, J. S., Schmidt, T. M., Falkow, S., and Tompkins, L. S. (1990). The agent of bacillary angiomatosis. *N. Engl. J. Med.* 323, 1573–1580. doi: 10.1056/NEJM199012063232301

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Riess, T., Anderson, B., Fackelmayer, A., Autenrieth, I. B., and Kempf, V. A. J. (2003). Rapid and efficient transposon mutagenesis of *Bartonella henselae* by transposome technology. *Gene* 313, 103–109. doi: 10.1016/S0378-1119(03)00636-X

Riess, T., Andersson, S. G. E., Lupas, A., Schaller, M., Schäfer, A., Kyme, P., et al. (2004). *Bartonella* adhesin A mediates a proangiogenic host cell response. *J. Exp. Med.* 200, 1267–1278. doi: 10.1084/jem.20040500

Riess, T., Dietrich, F., Schmidt, K. V., Kaiser, P. O., Schwarz, H., Schäfer, A., et al. (2008). Analysis of a novel insect cell culture medium-based growth medium for *Bartonella* species. *Appl. Environ. Microbiol.* 74, 5224–5227. doi: 10.1128/AEM.00621-08

Riess, T., Raddatz, G., Linke, D., Schäfer, A., and Kempf, V. A. J. (2007). Analysis of *Bartonella* adhesin A expression reveals differences between various *B. henselae* strains. *Infect. Immun.* 75, 35–43. doi: 10.1128/IAI.00963-06

Sander, A., Bühler, C., Pelz, K., von Cramm, E., and Bredt, W. (1997). Detection and identification of two *Bartonella henselae* variants in domestic cats in Germany. *J. Clin. Microbiol.* 35, 584–587. doi: 10.1128/JCM.35.3.584-587.1997

Schmid, M. C., Schülein, 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

Segers, F. H., Kešnerová, L., Kosoy, M., and Engel, P. (2017). Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *ISME J.* 11, 1232–1244. doi: 10.1038/ismej.2016.201

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951–960. doi: 10.1093/bioinformatics/bti125

Stahl, J., Bergmann, H., Göttig, S., Ebersberger, I., and Averhoff, B. (2015). *Acinetobacter baumannii* virulence is mediated by the concerted action of three phospholipases D. *PLoS One* 10:1–19. doi: 10.1371/journal.pone.0138360

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 14:473. doi: 10.1186/s12859-019-3019-7

Szczesny, P., and Lupas, A. (2008). Domain annotation of *trimeric autotransporter adhesin*s-daTAA. *Bioinformatics* 24, 1251–1256. doi: 10.1093/bioinformatics/btn118

Szczesny, P., Linke, D., Ursinus, A., Bär, K., Schwarz, H., Riess, T. M., et al. (2008). Structure of the head of the *Bartonella* adhesin BadA. *PLoS Pathog.* 4:e1000119. doi: 10.1371/journal.ppat.1000119

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., et al. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* 47, 10994–11006. doi: 10.1093/nar/gkz841

Tu, N., Carroll, R. K., Weiss, A., Shaw, L. N., Nicolas, G., Thomas, S., et al. (2017). A family of genus-specific RNAs in tandem with DNA-binding proteins control expression of the *badA* major virulence factor gene in *Bartonella henselae*. *Microbiologyopen* 6, 1–14. doi: 10.1002/mbo3.420

Tu, N., Lima, A., Bandeali, Z., and Anderson, B. (2016). Characterization of the general stress response in *Bartonella henselae*. *Microb. Pathog.* 92, 1–10. doi: 10.1016/j.micpath.2015.12.010

Vaca, D. J., Thibau, A., Schütz, M., Kraiczy, P., Happonen, L., Malmström, J., et al. (2019). Interaction with the host: the role of fibronectin and extracellular matrix proteins in the adhesion of Gram-negative bacteria. *Med. Microbiol. Immunol.* 2019:3. doi: 10.1007/s00430-019-00644-3

Wagner, A., and Dehio, C. (2019). Role of distinct type-IV-secretion systems and secreted effector sets in host adaptation by pathogenic *Bartonella* species. *Cell. Microbiol.* 21, 1–9. doi: 10.1111/cmi.13004

Weidensdorfer, M., Chae, J. I., Makobe, C., Stahl, J., Averhoff, B., Müller, V., et al. (2016). Analysis of endothelial adherence of *Bartonella henselae* and *Acinetobacter baumannii* using a dynamic human *ex vivo* infection model. *Infect. Immun.* 84, 711–722. doi: 10.1128/IAI.01502-15

Weidensdorfer, M., Ishikawa, M., Hori, K., Linke, D., Djahanschiri, B., Iruegas, R., et al. (2019). The *Acinetobacter trimeric autotransporter adhesin* Ata controls key virulence traits of *Acinetobacter baumannii*. *Virulence* 10, 68–81. doi: 10.1080/21505594.2018.1558693

Welch, D. F., Pickett, D. A., Slater, L. N., Steigerwalt, A. G., and Brenner, D. J. (1992). *Rochalimaea henselae* sp. nov., a cause of septicemia, bacillary angiomatosis, and parenchymal bacillary peliosis. *J. Clin. Microbiol.* 30, 275–280. doi: 10.1128/JCM.30.2.275-280.1992

Zbinden, R., Höchli, M., and Nadal, D. (1995). Intracellular location of *Bartonella henselae* cocultivated with Vero cells and used for an indirect fluorescent-antibody test. *Clin. Diagn. Lab. Immunol.* 2, 693–695. doi: 10.1128/cdli.2.6.693-695.1995

Zbinden, R., Michael, N., Sekulovski, M., Von Graevenitz, A., and Nadal, D. (1997). Evaluation of commercial slides for detection of immunoglobulin G against *Bartonella henselae* by indirect immunofluorescence. *Eur. J. Clin. Microbiol. Infect. Dis.* 16, 648–652. doi: 10.1007/BF01708554

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fmicb.2022.826829

Integrative Assessments on Molecular Taxonomy of *Acidiferrobacter thiooxydans* ZJ and Its Environmental Adaptation Based on Mobile Genetic Elements

Liyuan Ma[1,2], Weiyi Yang[1], Shanshan Huang[2], Rui Liu[1], Huiying Li[2], Xinping Huang[1], Junming Xiong[1] and Xueduan Liu[2]**

[1] Hubei Key Laboratory of Yangtze Catchment Environmental Aquatic Science, School of Environmental Studies, China University of Geosciences, Wuhan, China, [2] Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Daniel Yero,
Universidad Autónoma de Barcelona, Spain

Reviewed by:
Juan P. Cardenas,
Universidad Mayor, Chile
Francisco Luciano Issotta,
Fundación Ciencia & Vida, Chile

***Correspondence:**
Liyuan Ma
maly@cug.edu.cn
Xueduan Liu
xueduanliu@csu.edu.cn

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

Received: 01 December 2021
Accepted: 07 January 2022
Published: 16 February 2022

Citation:
Ma L, Yang W, Huang S, Liu R, Li H, Huang X, Xiong J and Liu X (2022) Integrative Assessments on Molecular Taxonomy of Acidiferrobacter thiooxydans ZJ and Its Environmental Adaptation Based on Mobile Genetic Elements. Front. Microbiol. 13:826829. doi: 10.3389/fmicb.2022.826829

Acidiferrobacter spp. are facultatively anaerobic acidophiles that belong to a distinctive *Acidiferrobacteraceae* family, which are similar to *Ectothiorhodospiraceae* phylogenetically, and are closely related to *Acidithiobacillia* class/subdivision physiologically. The limited genome information has kept them from being studied on molecular taxonomy and environmental adaptation in depth. Herein, *Af. thiooxydans* ZJ was isolated from acid mine drainage (AMD), and the complete genome sequence was reported to scan its genetic constitution for taxonomic and adaptative feature exploration. The genome has a single chromosome of 3,302,271 base pairs (bp), with a GC content of 63.61%. The phylogenetic tree based on OrthoANI highlighted the unique position of *Af. thiooxydans* ZJ, which harbored more unique genes among the strains from *Ectothiorhodospiraceae* and *Acidithiobacillaceae* by pan-genome analysis. The diverse mobile genetic elements (MGEs), such as insertion sequence (IS), clustered regularly interspaced short palindromic repeat (CRISPR), prophage, and genomic island (GI), have been identified and characterized in *Af. thiooxydans* ZJ. The results showed that *Af. thiooxydans* ZJ may effectively resist the infection of foreign viruses and gain functional gene fragments or clusters to shape its own genome advantageously. This study will offer more evidence of the genomic plasticity and improve our understanding of evolutionary adaptation mechanisms to extreme AMD environment, which could expand the potential utilization of *Af. thiooxydans* ZJ as an iron and sulfur oxidizer in industrial bioleaching.

Keywords: *Acidiferrobacter*, complete genome, comparative genomics, genome plasticity, environmental adaptation

INTRODUCTION

The type strain *Acidiferrobacter thiooxydans* m-1 was initially reported as "*Thiobacillus ferrooxidans*" owing to its iron-oxidizing capability, which was isolated from coal spoil refuse in Missouri, United States, over 35 years ago (Harrison, 1982). The representative genus *Acidithiobacillus* (formerly *Thiobacillus*) was the most widespread one with higher Fe/S oxidation ability, such as *A. ferrooxidans*, *A. ferrivorans*, *A. caldus*, and *A. thiooxidans* (Ma et al., 2021). However, the strain *Af. thiooxydans* m-1 harbored a higher chromosomal GC content and the analysis of partial 16S rRNA sequences revealed that it was quite distantly related to other acidophilic bacteria. This strain remained poorly studied on the physiology and phylogeny until 2011. Researchers renamed the strain as *Af. thiooxydans* and confirmed that it was closely related with the members of the family *Ectothiorhodospiraceae*, order *Chromatiales*, also including typically alkaliphilic and halophilic bacteria that could not oxidize iron and sulfur (Hallberg et al., 2011). Another member within the genus, *Acidiferrobacter* sp. strain SP-III/3 (DSM 27195), emerged as a new clade from the 16S rRNA-based phylogeny, which was isolated from AMD in Cartagena (Murcia, Spain) (Mitchell et al., 2004). The energy metabolism strategies from this genus, such as iron and sulfur oxidation features, were deeply analyzed *via* comparative genomics (Issotta et al., 2018). However, the calculated ANI relatedness between SPIII/3 and m-1 was lower than 95%, implying that they could not be assigned to the same species obviously.

In 2015, a novel chemolithoautotrophic sulfur oxidizer, called *Sulfurifustis variabilis* skN76, was isolated and characterized from the sediment of Lake Mizugaki (Kojima et al., 2015). It was closely related to *Af. thiooxydans*, but was distinct from other members within the family *Ectothiorhodospiraceae*. They were assigned into a novel family *Acidiferrobacteraceae* affiliated to the order *Acidiferrobacterales*. A year later, a similar novel sulfur oxidizer *Sulfuricaulis limicola* HA5, which was isolated from the sediment of Lake Harutori, was also established and assigned to the family *Acidiferrobacteraceae* (Kojima et al., 2016). Redundancies of the sulfite reduction and oxidation genes were observed in both of the two complete genomes, as represented by multiple copies of *dsrAB* and *aprAB* (Umezawa et al., 2016). The members of the family *Acidiferrobacteraceae* have been frequently detected in a variety of environments (Dyksma et al., 2016; Thouin et al., 2019; Wu et al., 2021), but only two assemblies of *Af. thiooxydans* were observed in the NCBI database. The restricted genome information limited the access to their genomic features in depth, which could give a deep understanding about their environmental adaptations.

Bioinformatic analyses highlighted the potential roles of various integrative mobile genetic elements (MGEs) in promoting the evolutionary adaptation of bacteria (Leal et al., 2020). Among the different types of MGEs, insertion sequences (ISs) contained a single open reading frame (ORF) encoding the transposase and inverted repeats at both ends. ISs transposition directly brought about gene deletion, inversion, and even rearrangement, which impacted a variety of bacterial life processes such as drug resistance, virulence, and metabolism (Vandecraen et al., 2017). The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas (CRISPR-associated proteins) system could integrate fragments of foreign nucleic acids into CRISPR arrays and embed in microbial genomes, achieving immunity for virus defense purposes (Faure et al., 2019). Prophages, namely, integrated virus genomes, were also common MGEs, which might be conducive to recruiting novel functionalities (Ramisetty and Sudhakari, 2019). Genomic islands (GIs) were mostly derived from horizontal gene transfer (HGT), which were the large fragments that enriched functional genes (Dobrindt et al., 2004). It was apparent that various MGEs played crucial roles in bacterial genome evolution and adaptation to specific environmental stress.

We isolated the strain *Af. thiooxydans* ZJ from Zijinshan Copper Mine, Fujian Province, China. The complete genome record CP080624 has been uploaded to NCBI that refreshed its draft genome (MDCF00000000.1). This study is the first time to report the complete genome of *Af. thiooxydans*. Experimental results have revealed that ZJ could adapt to a wide range of pH and temperature. However, there were many unknowns about the important MGEs that assisted the unique *Acidiferrobacteraceae* to acclimatize to the extreme acid environment. Further evidence of the adaptative mechanisms was urgently needed, especially supported by a complete genome in depth. In this study, the unique molecular taxonomy of *Af. thiooxydans* ZJ was confirmed by using orthoANI algorithm and pan-genome analysis. The MGEs in the single chromosome, such as IS, CRISPR, prophage, and GI, were identified to access its acidic environmental adaptation.

MATERIALS AND METHODS

Isolation and Cultivation

Acidiferrobacter thiooxydans ZJ was isolated from the AMD sample of Zijinshan Copper Mine (25°10′41″ N ∼ 25°11′44″ N, 116°24′00″ E ∼ 116°25′22″ E), China. The liquid samples from effusion pool were inoculated in 9K medium (pH 1.6) supplemented with 22.4 g/L ferrous sulfate and 5 g/L elemental sulfur. They were cultivated at 40°C in a rotary shaker at 170 rpm. The 9K medium contained the following components (g/L): $(NH_4)_2SO_4$ (3), K_2HPO_4 (0.5), KCl (0.1), $Ca(NO_3)_2$ (0.01), and $MgSO_4{\cdot}7H_2O$ (0.5). Strains were purified by serial gradient dilution and identified by 16S rRNA amplification and sequencing. According to the result from NCBI nucleotide collection (nr/nt) blast, the type strains provided by LPSN (List of Prokaryotic names with Standing in Nomenclature) (Parte, 2018) were employed to construct the phylogenetic tree using the maximum likelihood method by Molecular Evolutionary Genetics Analysis 11 (MEGA, v11) (Tamura et al., 2021). Bootstrap analysis was carried out on 1,000 replicate input data sets.

DNA Extraction and Sequencing

Cells of *Af. thiooxydans* ZJ were harvested by centrifugation (12,000 *g* for 10 min at 4°C) after cultivating to late exponential phase. Total DNA was extracted using the QIAamp DNA mini

kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Purified genomic DNA was quantified by a TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA, United States). The DNA with high quality (OD260/280 = 1.8–2.0, > 10 μg) was sequenced with a combination of PacBio RS and Illumina HiSeq platform supported by Shanghai Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China).

Genome Assembly and Annotation

The reads generated from PacBio and Illumina platform were filtered to form clean data. Then, the clean data were assembled into a contig using the hierarchical genome assembly process (HGAP) and Canu with default parameters (Chin et al., 2013; Koren et al., 2017). The last circular step was checked and finished manually, generating a complete genome of seamless chromosome. Finally, error correction of the PacBio assembly results was performed using the Illumina reads using Pilon (Walker et al., 2014). All of the analyses were performed using the free online platform of Majorbio Cloud Platform[1].

Identification of ORFs was conducted using Glimmer v3.02[2] (Delcher et al., 2007). ORFs less than 300 bp were discarded. Then, the remaining ORFs were queried against the databases of NCBI-nr, Swiss-Prot[3], KEGG[4], and COG[5] to do functional annotation by blastp 2.2.28 +. Moreover, the genome annotation was performed using the Rapid Annotation Subsystem Technology (RAST) server[6] (Aziz et al., 2008). In addition, tRNAs were identified by the tRNAscan-SE (v1.23[7]) (Chan et al., 2021) and rRNAs were determined by the Barrnap v0.9[8].

Comparative Genomics Analysis

Based on the 16S rRNA phylogenetic tree, the complete genomes of *Sulfurifustis variabilis* skN76, *Sulfuricaulis limicola* HA5, *Ectothiorhodospira* sp. BSL-9, and *Thioalkalivibrio sulfidiphilus* HL-EbGr7 from *Acidiferrobacteraceae* and *Ectothiorhodospiraceae*, together with the complete genomes of *Acidithiobacillus ferrooxidans* ATCC 23270, *Acidithiobacillus ferrivorans* SS3, *Acidithiobacillus thiooxidans* ATCC 19377, and *Acidithiobacillus caldus* ATCC 51756 from *Acidithiobacillaceae*, were selected for comparative studies.

OrthoANI values among the strains from *Acidiferrobacter* and the two groups mentioned above were calculated based on whole-genome sequences. Heatmaps were plotted on the OAT: OrthoANI Tool (v0.93.1) subsequently (Lee et al., 2016). Average Amino acid Identity (AAI) was calculated on the online site[9] (Rodriguez-R and Konstantinidis, 2014).

The genome sequences of the strains from the two groups were annotated with Prodigal (Hyatt et al., 2010). Based on the Bacterial Pan-genome Analysis tool (BPGA v1.3), the pan-genome and core-genome were estimated using the USEARCH program (v11.0.667) available in BPGA, with a 50% cutoff of sequence identity (Chaudhari et al., 2016).

Mobile Genetic Elements Identification

Insertion sequence and transposases distributed over *Af. thiooxydans* ZJ genome were predicted and classified using the ISfinder platform[10] based on online blast analysis with an E-value of 1e-10 (Siguier et al., 2006). The information about the position, family type, and copy number of all the ISs was obtained from the blast results. Functional regions around the ISs were identified and selected to generate a physical gene map.

The CRISPRCasFinder web tool[11] (Couvin et al., 2018) was employed to identify the CRISPR/Cas array, including the repeat sequences and spacer sequences. The differences of the repeat sequences at each base site were visualized using WebLogo[12] online server (Crooks et al., 2004). The RNA secondary structure of the repeat sequence was predicted by RNAfold[13] (Mathews et al., 1999). The spacers were further annotated by blastn against the NCBI-nt database.

Integrated bacteriophages (prophages) prediction and the active evaluation were conducted using Prophage Hunter[14] (Song et al., 2019). There are two alternative strategies: either incorporating similarity searches to increase accuracy or skipping it to raise the possibility of finding novel phages. In order to ensure the accuracy, we did not ignore the similarity matching in this study. Then, the qualified phages were classified taxonomically at the family and genus levels. The genome lengths of prophages from different families were calculated.

The GI was predicted by Island-viewer 4 online server[15] (Bertelli et al., 2017) joining three methods: IslandPath-DIMOB (Hsiao et al., 2005), IslandPick (Langille et al., 2008), and SIGI-HMM (Waack et al., 2006). The proteins corresponding to GI predicted by at least one way from Island-viewer 4 online server were counted.

RESULTS AND DISCUSSION

Genomic Features of *Acidiferrobacter thiooxydans* ZJ

After filtering, data from PacBio and Illumina corresponded to an 81 × and 132 × coverage. The circular genome map of *Af. thiooxydans* ZJ is shown in **Figure 1A**. The genome has a single chromosome of 3,302,271 bp, with a GC content of 63.61%. There were 3,343 coding DNA sequences (CDSs), 46 tRNAs, and 3 rRNAs in the complete genome. The average gene length was 850 bp and the coding density (%) was 87.33%. In addition, 56 repeated sequences were observed with a total

[1]www.majorbio.com
[2]http://ccb.jhu.edu/software/glimmer/index.shtml
[3]http://uniprot.org
[4]http://www.genome.jp/kegg/
[5]http://www.ncbi.nlm.nih.gov/COG
[6]http://rast.nmpdr.org
[7]http://lowelab.ucsc.edu/tRNAscan-SE
[8]github.com/tseemann/barrnap
[9]http://enve-omics.ce.gatech.edu/aai/
[10]https://www-is.biotoul.fr/
[11]https://crisprcas.i2bc.paris-saclay.fr
[12]http://weblogo.threeplusone.com/
[13]http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi
[14]https://pro-hunter.bgi.com/
[15]https://www.pathogenomics.sfu.ca/islandviewer/

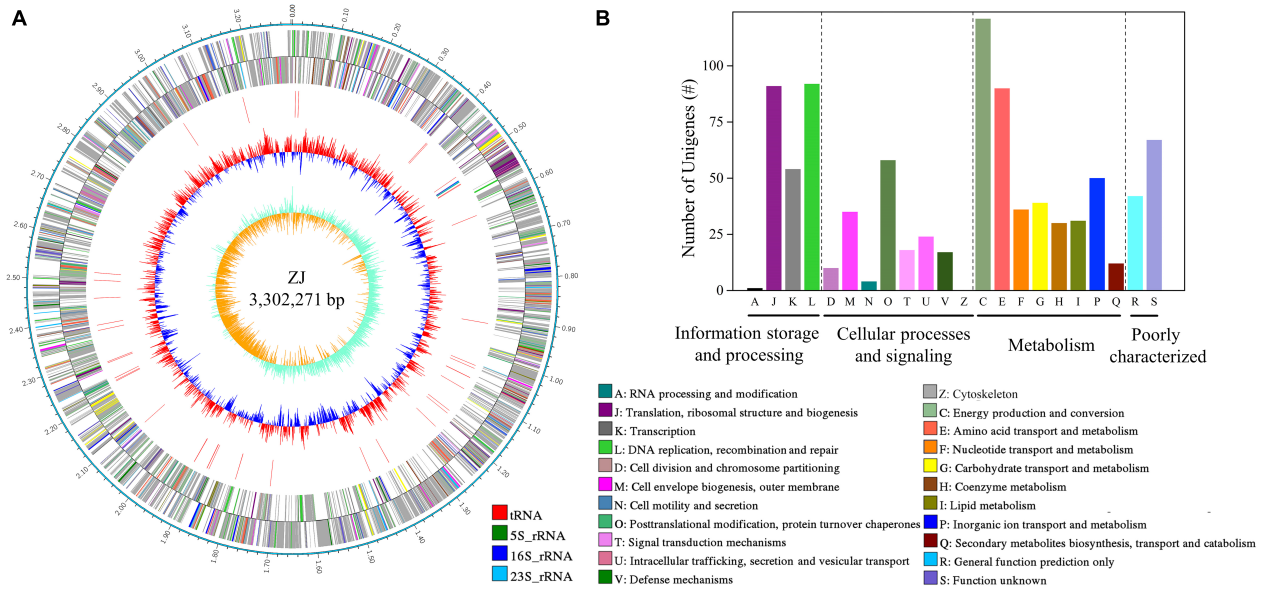


FIGURE 1 | Circular map of *Af. thiooxydans* ZJ **(A)** and the classification of COG functions **(B)**. From outside to inside: rules for reads; CDS on sense DNA strand; CDS on antisense DNA strand; rRNA and tRNA; GC content; GC skew.

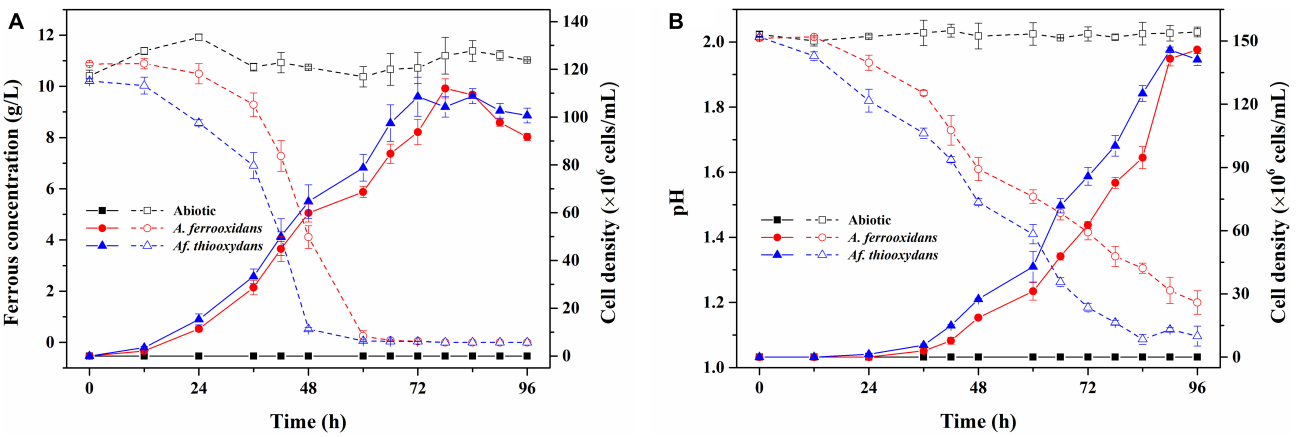


FIGURE 2 | The growth and iron/sulfur oxidation rate of *Af. thiooxydans* ZJ and *A. ferrooxidans* ATCC 23270 under the same culture conditions (pH 2.0, Temperature 30°C). The dashed lines represent the ferrous concentration in the left *y*-axis and the solid lines represent the cell density in the right *y*-axis **(A)**. The dashed lines represent the pH in the left *y*-axis and the solid lines represent the cell density in the right *y*-axis **(B)**.

of 11,206 bp in length. There were 2,667 genes uncovered by NCBI-nr, KEGG, Swiss-Prot, GO, and COG databases, accounting for 78.65% of all the predicted genes. The COG annotation (**Figure 1B**) revealed that the genes belonging to C (energy production and conversion), E (amino acid transport and metabolism), J (translation, ribosomal structure, and biogenesis), and L (replication, recombination, and repair) ranked as the top four COG classifications.

It has been reported that *Af. thiooxydans* could obtain energy by oxidizing ferrous iron, pyrite, element sulfur, sulfide, or tetrathionate in the oligotrophic habitat (Hallberg et al., 2011). As an efficient iron and sulfur oxidizer, *Af. thiooxydans* ZJ showed faster growth and oxidation rate of iron and sulfur than *A. ferrooxidans* ATCC 23270 under the same culture conditions (pH 2.0, Temperature 30°C, **Figure 2**), which was an essential individual of the microbial community in the bioleaching system (Wu et al., 2021). Herein, the genes encoding blue copper-containing protein rusticyanin and the sulfur-oxidizing gene cluster (*sox*) were identified in *Af. thiooxydans* ZJ genome, supporting the strong iron and sulfur oxidation capacity (**Figure 2**). Besides, the *recA*-dependent acid-tolerance system (COG0468), which was an inducible pathway involved in DNA repair, has been identified in *Af. thiooxydans* ZJ, indicating that *Af. thiooxydans* ZJ may present a flexible response and strong adaptation to the extremely acid environments based on recombination and repair, just like the common

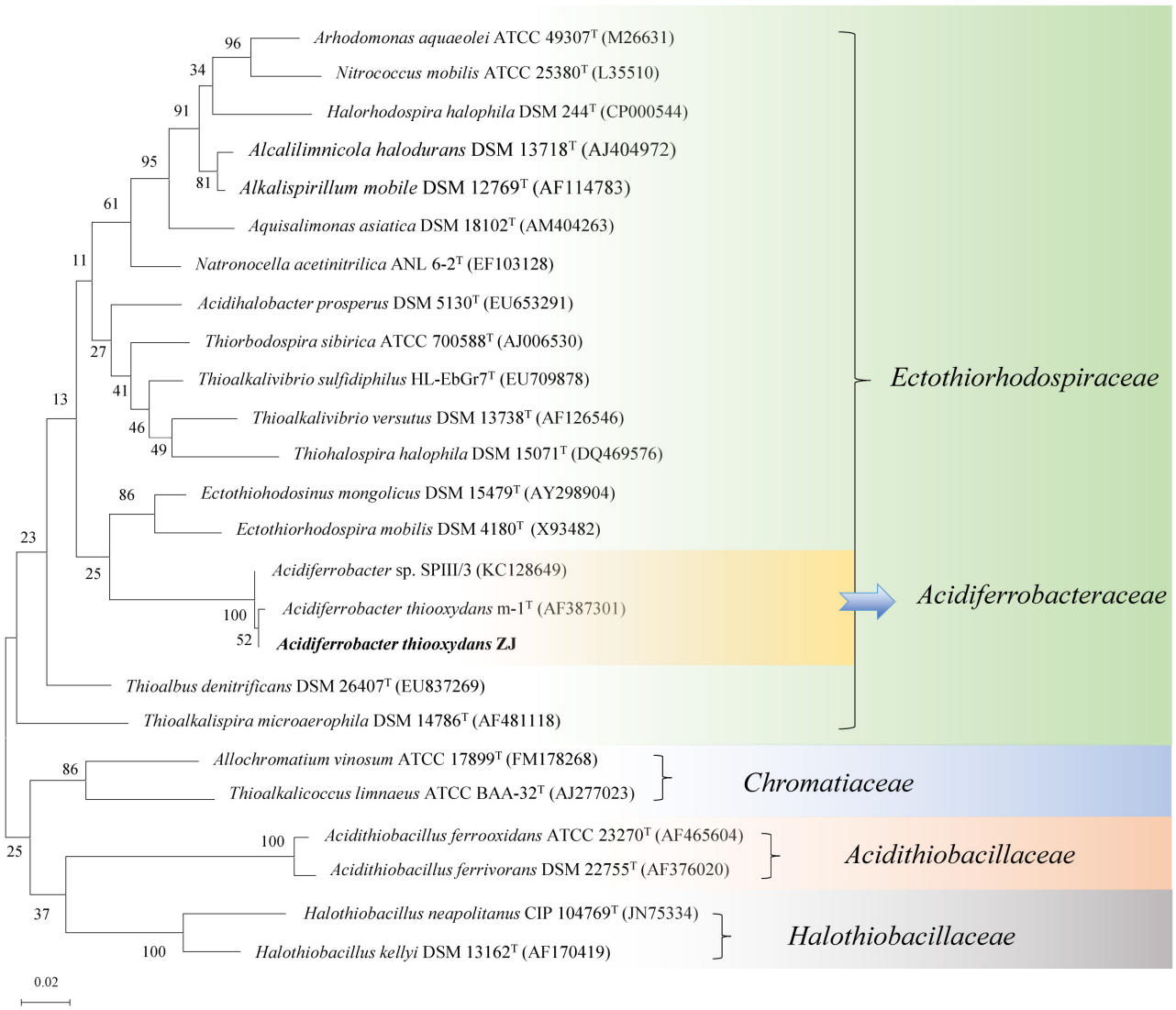


FIGURE 3 | Phylogenetic tree of *Af. thiooxydans* ZJ with the neighboring standard species based on 16S rRNA sequences. The tree was constructed by maximum likelihood method with bootstrap values calculated from 1,000 trees. The numbers at each clustering node indicate the percentage of bootstrap supporting. The scale bar 0.02 indicates evolutionary distance.

acidophiles in mining areas (Bellenberg et al., 2019; Ma et al., 2019).

Phylogenetic Position of *Acidiferrobacter thiooxydans* ZJ Based on Comparative Genomics

According to the blast result of the NCBI-nr database, the nearby standard strains (25 strains) from LPSN were selected to construct a phylogenetic tree based on their 16S rRNA genes. As shown in **Figure 3**, the different families (*-aceae*) were well separated within the tree, and *Af. thiooxydans* ZJ formed a clade with *Af. thiooxydans* m-1 and *Acidiferrobacter* sp. SPIII/3. In addition, they were more closely related to *Ectothiorhodospira* and *Thioalkalivibrio* branches. However, they departed from the outgroup cluster formed by *A. ferrooxidans* and *A. ferrivorans*, which belonged to *Acidithiobacillaceae*. Only a few bacteria from the family *Ectothiorhodospiraceae* have been identified as iron oxidizers (Davis-Belmar et al., 2008; Hallberg et al., 2011). One was *Thiodictyon* sp. strain L7, which oxidized ferrous iron in anaerobic conditions; the other was *Acidihalobacter prosperus* (originally as *Thiobacillus prosperus*), which was capable of oxidizing both iron and sulfur compounds (Huber and Stetter, 1989; Pablo Cárdenas et al., 2015). This oxidation character also supported a separate status of *Acidithiobacillaceae* from *Ectothiorhodospiraceae*.

Based on the 16S rRNA phylogenetic analysis results, we further calculated the OrthoANI distance among the nearby complete genomes of 11 strains. The genomic features of the selected genomes are summarized in **Table 1**. As shown in

TABLE 1 | Overview of the complete genomes selected from *Ectothiorhodospiraceae* and *Acidithiobacillaceae*.

Species	Isolation source	Accession number	Assembly level	Genome size/Mb	GC%	Gene	Protein	rRNA	tRNA	References
Acidiferrobacter thiooxydans ZJ	Zijinshan Copper Mine	CP080624.1	Complete	3.15	63.6	3,115	3,039	3	46	This study
Acidiferrobacter thiooxydans m-1	Coal strip mine refuse	NZ_PSYR00000000.1	Contig	3.25	63.7	3,124	3,037	3	45	Issotta et al., 2018
Acidiferrobacter sp. SPIII/3	Mining area near to La Esperanza in Murcia, Spain	NZ_CP027663.1	Complete	3.40	64.2	3,414	3,295	3	45	Issotta et al., 2018
*Sulfurifustis variabili*s skN76	Sediment of a freshwater lake	NZ_AP014936.1	Complete	3.96	67.5	3,909	3,844	3	46	Umezawa et al., 2016
Sulfuricaulis limicola HA5	Sediment from a meromictic lake from residential area	NZ_AP014879.1	Complete	2.86	61.4	2,803	2,742	3	46	Umezawa et al., 2016
Ectothiorhodospira sp. BSL-9	Big Soda Lake, Nevada	NZ_AP014879.1	Complete	3.55	63	3,300	3,158	6	46	Hernandez-Maldonado et al., 2016
Thioalkalivibrio sulfidiphlius HL-EbGr7	Thiopaq bioreactor used to remove H_2S from biogas	NC_011901.1	Complete	3.46	65.1	3,372	3,266	3	44	Muyzer et al., 2011
Acidithiobacillus ferrooxidans ATCC 23270	Acid, bituminous coal mine effluent	NC_011761.1	Complete	2.98	58.8	3,075	2,893	6	82	Valdés et al., 2008
Acidithiobacillus ferrivorans SS3	Sediment, Norilsk, Russia	NC_015942.1	Complete	3.21	56.6	3,200	3,038	6	47	Liljeqvist et al., 2011
Acidithiobacillus thiooxidans ATCC 19377	Kimmeridge clay	NZ_CP045571.1	Complete	3.42	53	3,573	3,388	6	76	Valdes et al., 2011
Acidithiobacillus caldus ATCC 51756	Coal spoil at Kingsbury Mine	NZ_CP026328.2	Complete	2.73	61.7	2,706	2,550	6	47	Tapia et al., 2012

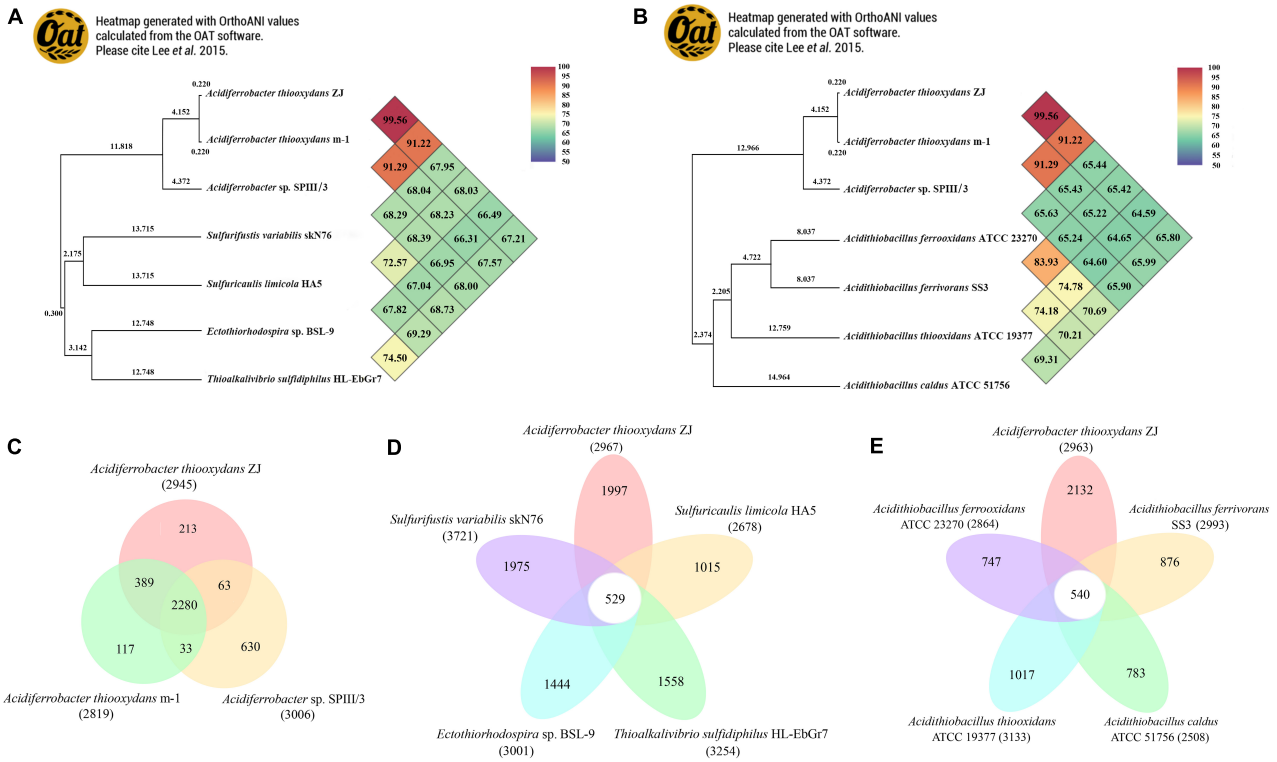


FIGURE 4 | Heatmap generated with OrthoANI values calculated using the OAT software between *Acidiferrobacter* and other closely related species from *Ectothiorhodospiraceae* **(A)** and *Acidithiobacillaceae* **(B)**. The numbers at each clustering node indicate the branch length. Pan-Genome analysis among *Acidiferrobacter* **(C)**, *Af. thiooxydans* ZJ with *Ectothiorhodospiraceae* **(D)**, and *Acidithiobacillaceae* **(E)**.

Figure 4, the strain *Af. thiooxydans* ZJ and *Af. thiooxydans* m-1 belonged to the same genomic species (OrthoANI value = 99.56%), but they were distinct from the genomic species of SPIII/3 (OrthoANI value = 91.22% and 91.29%). An OrthoANI value of 95–96% could be used as a cutoff for bacterial species demarcation, which corresponded to a cutoff point of 70% digital DNA–DNA hybridization (DDH) (Lee et al., 2016). The strains within the family maintained distant relationships with the similarity as low as 66.49–68.03% (**Figure 4A**). Recent studies have reassigned *Af. thiooxydans* to the new *Acidiferrobacteraceae* family with *Sulfurifustis variabilis* skN76 (Kojima et al., 2015) and *Sulfuricaulis limicola* HA5 (Kojima et al., 2016), but the OrthoANI between them were not higher than that with other members from *Ectothiorhodospiraceae*, such as *Ectothiorhodospira* sp. BSL-9 and *Thioalkalivibrio sulfidiphilus* HL-EbGr7. Meantime, as expected, the lower genomic relatedness with *Acidithiobacillus* was also confirmed by their greater distances with all the similarity less than 66% (**Figure 4B**). In addition, the calculation results of AAI (lower than 60%, **Supplementary Table 1**) also demonstrated that the genus *Acidiferrobacter* occupied a unique taxonomic position, which was distant from *Ectothiorhodospiraceae* and *Acidithiobacillaceae*.

Pan-genome analysis was performed to identify the corresponding core and dispensable genome. A core genome containing 2,280 CDSs was observed within *Acidiferrobacter*; meantime, *Af. thiooxydans* ZJ and m-1 shared 389 genes, while SPIII/3 harbored more unique genes (**Figure 4C**). Compared with the genomes from the broad *Ectothiorhodospiraceae*, *Af. thiooxydans* ZJ harbored 1,997 unique genes apart from other genomes (**Figure 4D**), implying a relatively high degree of genomic diversity and low relatedness with the other members of the family mentioned above. Meanwhile, there were 540 genes shared with *Acidithiobacillus*, and the strain-specific genes in *Af. thiooxydans* ZJ were as high as 2,132 among the selected *Acidithiobacillus* species (**Figure 4E**). Previous studies have demonstrated that the genetic traits concerning adaptation, resistance and virulence were more often governed by dispensable genome (Maistrenko et al., 2020); therefore, *Af. thiooxydans* ZJ was hypothetically conferred selective advantages and special adaptability to the similar extreme environment.

Insertion Sequences and Transposases in *Acidiferrobacter thiooxydans* ZJ

Insertion sequences are the simplest type of MGEs in microbial genomes, which could be classified based on general characteristics of their DNA sequences and relevant transposases (Vandecraen et al., 2017). There were 12 types of ISs in *Af. thiooxydans* ZJ genome, belonging to 5 families with a total of 46 copies. The IS200/IS605 family, including ISAfe8, ISAba30, ISMex34, ISSoc9, and ISCARN6, exhibited the highest copy

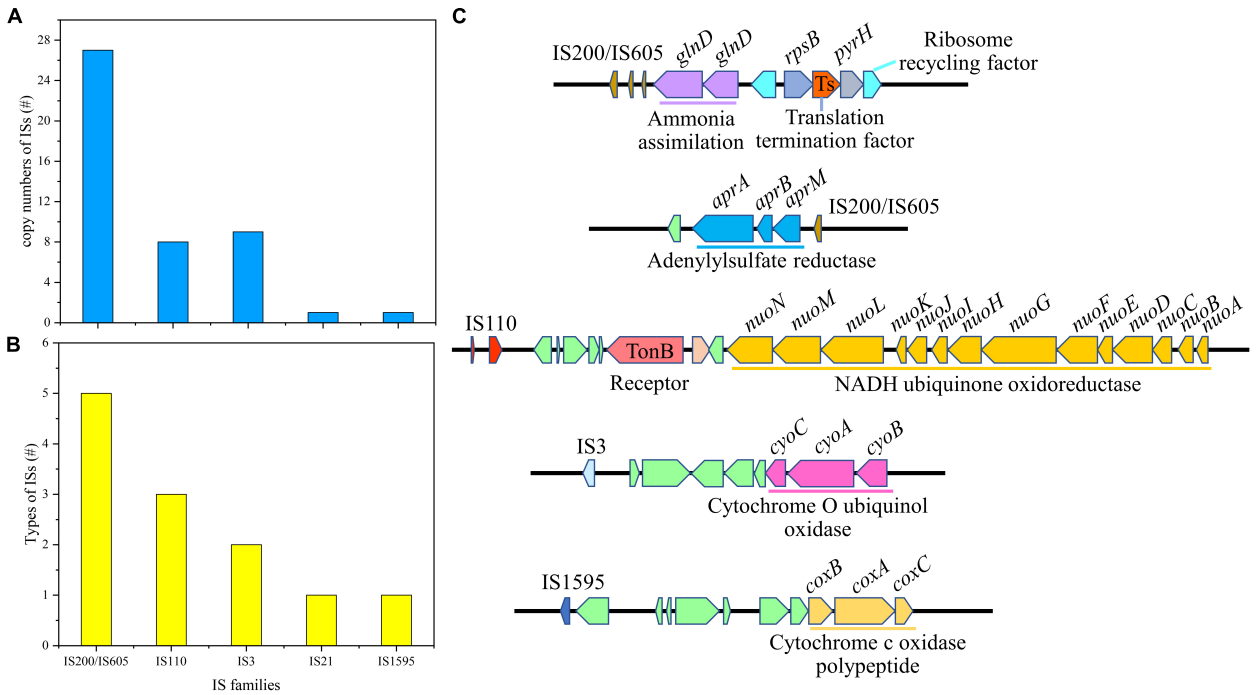


FIGURE 5 | The copy number of insertion sequences **(A)**, the families of ISs **(B)**, and the putative laterally transferred regions contain ISs **(C)**.

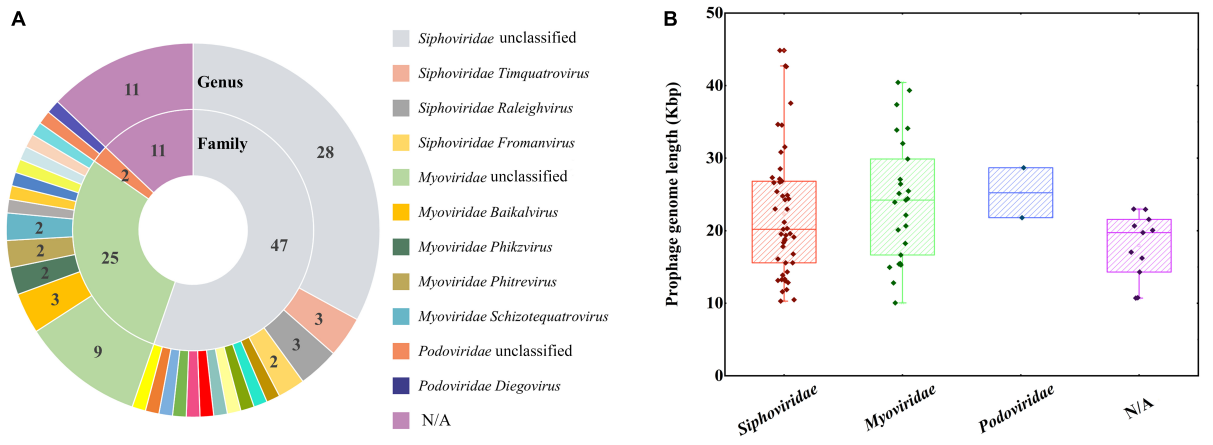


FIGURE 6 | The closest phages classified according to family and genus levels **(A)**. The numbers represented the closest phages belonging to the different families. A boxplot of prophage genome lengths based on closest phages family taxonomy **(B)**.

number of 27 in the genome (**Figures 5A,B**). It has been reported that IS200/IS605 was an ancient and stable IS family with the smallest known DNA transposases, working based on the so-called peel-and-paste mechanism (Barabas et al., 2008). The transposases in *Af. thiooxydans* ZJ genome could be categorized into three types, belonging to DEDD, DDE, and Y1 HUH (**Supplementary Table 2**). Unlike DDE ISs, ISs with DEDD or Y1 HUH transposases neither carry the terminal inverted repeats (IRs) nor generate flanking direct repeats (DRs) on insertion; meanwhile, Y1 HUH encoding ISs were able to form hairpin secondary structures particularly (Siguier et al., 2014). This secondary hairpin structure was formed to transposase specific recognition and catalyzed single-strand cleavage that could help stabilize the nuclear protein complex. We herein supposed that ISs of *Af. thiooxydans* ZJ may utilize a variety of strategies to shape the *Af. thiooxydans* ZJ genome effectively.

It was well acknowledged that ISs could mobilize neighborhood genes in certain cases. The adjacent gene compositions of ISs were examined in *Af. thiooxydans* ZJ. As shown in **Figure 5C**, IS200/IS605 was mainly located upstream of the gene (*glnD*) coding for ammonia assimilation along with the functional regions of translation termination and

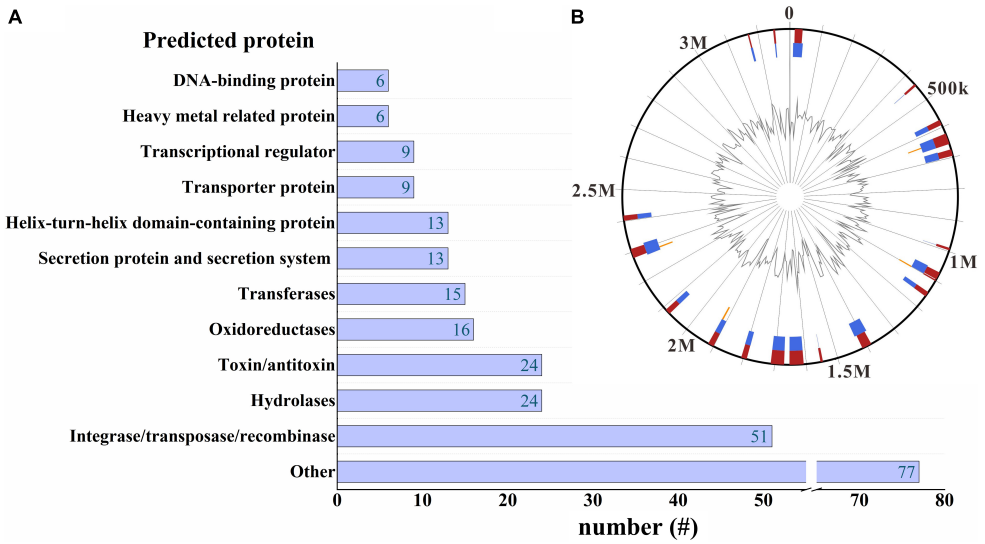


FIGURE 7 | A horizontal bar graph of the predicted proteins located on the GI regions **(A)**. GI distribution in *Af. thiooxydans* ZJ genome **(B)**. The blue and orange lines represent the GI predictions by IslandPath-DIMOB and SIGI-HMM, respectively, and the red line in the outer circle represents the comprehensive results by different methods. The peak map in the inner circle represents the variation of the GC content.

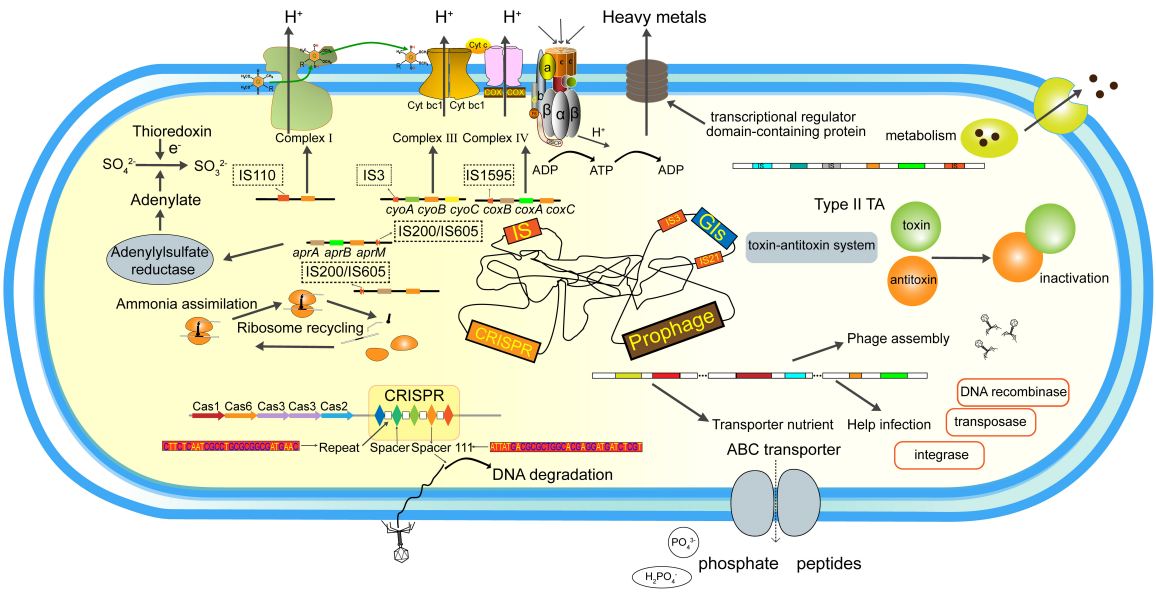


FIGURE 8 | Functional repertoires of the *Af. thiooxydans* ZJ. The MGEs and their predicted gene products identified in this study.

ribosome recycling. IS200/IS605 was also located downstream of the genes (*aprA*, *aprB*, and *aprM*) coding adenylylsulfate reductase. Adenylylsulfate reductase mainly utilized adenylate as a substrate and thioredoxin as an electron donor to reduce active sulfate to sulfite (Zhang et al., 2016). IS110 was flanked by the genes coding for NADH ubiquinone oxidoreductase, which was involved in electron transfer in the respiratory system. Besides, IS3 presented upstream of the genes (*cyoA*, *cyoB*, and *cyoC*) coding for cytochrome O ubiquinone oxidase, and IS1595 was located upstream of the genes (*coxB*, *coxA*, and *coxC*) coding for cytochrome C oxidase polypeptide (**Figure 5C**). All these enzymes are important oxidases involved in bacterial electron transfer (Dinamarca et al., 2002), providing essential functions for *Af. thiooxydans* ZJ to survive in the oligotrophic and stressful environment caused by heavy metals. Hence, the functional genes around these ISs may contribute to the processing of sulfide, assimilation of ammonia from the surroundings, and electron transfer, improving the favorable survival and strong adaptation of *Af. thiooxydans* ZJ in AMD environment.

CRISPR/Cas Systems in *Acidiferrobacter thiooxydans* ZJ

The CRISPR/Cas systems were adaptive immunity systems that served as viral and plasmid defense mechanisms developed by bacteria and archaea (Faure et al., 2019). The *Af. thiooxydans* ZJ genome harbored a quite large CRISPR array (13,779 bp), which contained 230 direct repeats (28 bp each) separated by 229 different spacers of similar sizes (32–36 bp) (**Supplementary Table 3**), suggesting that it was fairly defensive to the foreign genetic elements along the evolutionary process (Kojima et al., 2015). Based on the *cas* gene arrangement and the presence of Cas3 (Cas1-Cas6-Csy3-Cas3-Cas2), the CRISPR/Cas system of *Af. thiooxydans* ZJ could be classified as subtype I-F; additionally, type I CRISPR/Cas systems were speculated to directly target DNA against invading viruses (Makarova et al., 2015). Interestingly, our results of blastn program against the NCBI-nt database showed that the spacer 111 (ATTATGACGCGCTGGCACGAGGATGATCTCGT) has a similarity of 100% to a fraction genome of the *Myoviridae* sp. isolate ct6gw13. *Myoviridae* has been reported to play a major part in viral populations in the sulfidic mine tailings (Gao et al., 2020), while further studies regarding viral diversity and functions in the acid mine environment were needed to expand the knowledge.

The formation of secondary structure of CRISPR repeats was closely related to its stability and mechanism. If the predicted secondary structure was mainly formed by "loop," the interaction between spacer and foreign DNA was performed by a single "repeat-spacer" unit during the transcription process (Kunin et al., 2007). The repeats may serve to mediate the coupling cleavage of exogenous genetic materials and Cas-encoded proteins. Otherwise, the contact between spacers and foreign DNA was assisted by two adjacent repeats (Kunin et al., 2007). As shown in **Supplementary Figure 1**, the 28-bp repeats in *Af. thiooxydans* ZJ were rather conservative. Only at positions 15, 16, and 19 did the site conservation fluctuate slightly. What is more, the stem length of the proposed RNA secondary structure of repeat sequence was 8 bp, that is to say, the stem-loop structure was dominated by "stem." Thus, it could be speculated that the defense function of the spacer against the foreign DNA invasion may be facilitated through the complementation of two adjacent repeats (Kunin et al., 2007).

In short, *Af. thiooxydans* ZJ has evolved a unique I-F CRISPR/Cas system, which may have resisted the intrusion of foreign viruses, attaching great significance for its adaptation and development in the extremely acidic environment.

Prophages in *Acidiferrobacter thiooxydans* ZJ

Bacteriophage, as a source of foreign DNA, are increasingly recognized to contribute to gene flow in prokaryotes, which also led to bacterial genome plasticity (Dobrindt et al., 2010). Prophages played prominent roles in the host-adaptive traits and genetic diversification by delivering functional genes among different strains (Ramisetty and Sudhakari, 2019). Based on the score of each candidate, regions with scores higher than 0.8 were defined as an "active" prophage; the score of an "ambiguous" one ranged from 0.5 to 0.8, and a score lower than 0.5 was defined as an "inactive" one (Song et al., 2019). In the genome of *Af. thiooxydans* ZJ, 85 prophage candidates were identified, of which three candidates were the ambiguous ones. None of these phages was predicted to be active (**Supplementary Table 4**). A total of 1,807 fragments were obtained, including CDS information and functional annotation results based on NCBI-nr, Pfam, and InterPro databases (**Supplementary Table 5**). These prophage genes were mainly related to encoding phage structure and assembly proteins (e.g., head–tail protein, tape measure protein), or the enzymes that play vital roles in the whole process of phage infection (e.g., transposase, site-specific DNA recombinase, and integrase). There was a major category of membrane-associated protein structures named ATP binding cassette (ABC) transporters, which are involved in the nutrient uptake through medium- and high-affinity pathways in bacteria (Locher, 2016). In this study, it was estimated to be mainly involved in the transport of phosphate or peptides. Meanwhile, genes encoding TonB-dependent receptors have been identified, which participated in the uptake of rare earth metals (Noinaj et al., 2010; Ochsner et al., 2019), membrane homeostasis (Menikpurage et al., 2019), and secretion of proteins (Gómez-Santos et al., 2019). Moreover, many genes encoding hydrolases, diguanylate cyclase, peptidases, and the toxin–antitoxin (TA) system have also been observed, which may improve the survival and growth of *Af. thiooxydans* ZJ (Costa et al., 2018).

The closest phage relatives were also identified, of which 74 prophage genome regions mainly belonged to the family *Siphoviridae* (47, with a percentage of 55.29%), *Myoviridae* (25, with a percentage of 29.41%), and *Podoviridae* (2, with a percentage of 2.35%) (**Figure 6A** and **Supplementary Table 6**). All the three families were affiliated to the order *Caudovirales*, which were the tailed double-stranded DNA bacteriophages, including 14 families and 927 genera (Lefkowitz et al., 2018). The length of prophage fragments was presented in **Figure 6B**. The result was consistent with the most common phages in nature (Ackermann, 2007).

Genomic Islands in *Acidiferrobacter thiooxydans* ZJ

GIs are commonly believed to be relics of HGT and clusters of genes encoding different functions, which plays a role in genome plasticity and evolution, offering a selective advantage for host bacteria (Dobrindt et al., 2004; Juhas et al., 2009). A total of 23 GIs were identified in the *Af. thiooxydans* ZJ genome ranging from 4.164 kb to 44.743 kb in size (**Figure 7B**). There were 415 predicted genes located in these GIs, among which 152 genes were annotated as "hypothetical proteins" with unknown functions. The annotated proteins were mostly related to integrase/transposase/recombinase (51 genes) and antitoxin/toxin (24 genes) (**Figure 7A**). In addition, some GIs also contained the genes encoding proteins for maintaining bacterial survival, metabolism, and growth (e.g., hydrolases, secretion system, and oxidoreductase). These strongly evidenced that GIs in *Af. thiooxydans* ZJ genome helped transfer a large

number of gene families to the host bacteria to improve its adaptability in high metal concentration environment.

Intriguingly, integrases, transposases, and recombinases were powerful tools for insertion mutations, which also played a critical role in shaping the genome and events leading to speciation (Kazazian, 2004). There are 29 genes in the GI region involved in encoding IS family transposases (IS3, IS21, ISL3, etc.), promoting the movement of DNA segments to new locations within and between genomes. The majority of the potential GIs were noted to flank by these genes, suggesting that transposases may be associated with the acquisition of GIs (Rao et al., 2020; **Supplementary Table 7**).

It was noteworthy that the TA system was widely distributed in prokaryotes, which play a wide range of biological functions, including plasmid addiction, antibiotic tolerance, and defense against phages (Otsuka, 2016; Harms et al., 2018). Many genes regarding bacterial toxicity have been found in the GIs of *Af. thiooxydans* ZJ, such as the genes encoding RelE/ParE family toxin, VapC family toxin, Phd/YefM family antitoxin, and addiction module protein. Most of them belonged to the type II TA system, which neutralized toxicity by forming a protein–protein complex between antitoxin and toxin (Unterholzner et al., 2013). Previous studies have demonstrated that the TA system was involved in the stabilization of large genomic segments (Szekeres et al., 2007) and its addictive properties allowed them to integrate stably in bacterial chromosomes (Leplae et al., 2011). We hypothesized that GIs might play an essential role in the evolution of virulence in *Af. thiooxydans* ZJ. Meanwhile, genes encoding heavy metal-responsive transcriptional regulator and heavy metal-binding domain-containing protein have also been identified in GIs. Combined with previous surveys of *Acidithiobacillus* (Hemme et al., 2016; Li et al., 2019), the influence of HGT on the evolution of heavy metal resistance in *Af. thiooxydans* ZJ surviving in extreme environments needed to be further investigated.

CONCLUSION

In the present study, the iron and sulfur oxidizer *Af. thiooxydans* ZJ was isolated from AMD, yielding a complete genome with 3,302,271 bp with a high GC content of 63.61%. Comparative genomic analysis revealed that *Af. thiooxydans* ZJ was distinctive within the *Acidiferrobacteraceae* family phylogenetically, while it was more closely related to *Acidithiobacillaceae* physiologically. Subsequently, a considerable various repertoire of MGEs were assessed, including 5 families of ISs, an I-F subtype CRISPR/Cas system, 85 prophage regions, and 23 GIs generally (**Figure 8**). ISs may harbor different insertion patterns based on transposases. Most GIs were also annotated to contain the genes encoding integrase/transposase/recombinase that facilitate the insertion and integration of external fragments. The blast results of CRISPR spacers suggested that it may resist phage intrusion, which was also supported by the discovery of prophages. The prophage regions encoded functional enzymes beneficial for the growth probably fought and derived by *Af. thiooxydans* ZJ. Collectively, genome plasticity regarding the gain/loss of MGEs contributed to the environmental adaptation of *Af. thiooxydans* ZJ in extremely acid niche with high metal concentrations. The effects of MGEs on bacterial genome shaping and adaptation mechanisms deserved further explorations extensively and deeply. Meanwhile, more experiments about the relevant genes are needed to verify the effectiveness.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

LM, WY, and XL: conceptualization. WY, SH, RL, and HL: methodology. XH and JX: validation. WY, XH, and JX: formal analysis. LM and WY: investigation and writing—original draft preparation. LM and XL: resources, supervision, and project administration. LM: data curation. SH, RL, HL, and XL: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (42007306 and 52174341) and the Provincial Water Resources Development Fund Project of Yangquan Municipal Water Resources Bureau (2019046477).

ACKNOWLEDGMENTS

We would like to thank Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China) for the sequencing support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.826829/full#supplementary-material

REFERENCES

Ackermann, H. W. (2007). 5500 Phages examined in the electron microscope. *Arch. Virol.* 152, 227–243. doi: 10.1007/s00705-006-0849-1

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

Barabas, O., Ronning, D. R., Guynet, C., Hickman, A. B., Ton-Hoang, B., Chandler, M., et al. (2008). Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell* 132, 208–220. doi: 10.1016/j.cell.2007.12.029

Bellenberg, S., Huynh, D., Poetsch, A., Sand, W., and Vera, M. (2019). Proteomics reveal enhanced oxidative stress responses and metabolic adaptation in *Acidithiobacillus ferrooxidans* biofilm cells on pyrite. *Front. Microbiol.* 10:592. doi: 10.3389/fmicb.2019.00592

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343

Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/nar/gkab688

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6:24373. doi: 10.1038/srep24373

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Costa, A. R., Monteiro, R., and Azeredo, J. (2018). Genomic analysis of *Acinetobacter baumannii* prophages reveals remarkable diversity and suggests profound impact on bacterial virulence and fitness. *Sci. Rep.* 8:15346. doi: 10.1038/s41598-018-33800-5

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

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

Davis-Belmar, C. S., Nicolle, J. L. C., and Norris, P. R. (2008). Ferrous iron oxidation and leaching of copper ore with halotolerant bacteria in ore columns. *Hydrometallurgy* 94, 144–147.

Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23, 673–679. doi: 10.1093/bioinformatics/btm009

Dinamarca, M. A., Ruiz-Manzano, A., and Rojo, F. (2002). Inactivation of cytochrome o ubiquinol oxidase relieves catabolic repression of the *Pseudomonas* putida GPo1 alkane degradation pathway. *J. Bacteriol.* 184, 3785–3793. doi: 10.1128/JB.184.14.3785-3793.2002

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414–424. doi: 10.1038/nrmicro884

Dobrindt, U., Zdziarski, J., Salvador, E., and Hacker, J. (2010). Bacterial genome plasticity and its impact on adaptation during persistent infection. *Int. J. Med. Microbiol.* 300, 363–366. doi: 10.1016/j.ijmm.2010.04.010

Dyksma, S., Bischof, K., Fuchs, B. M., Hoffmann, K., Meier, D., Meyerdierks, A., et al. (2016). Ubiquitous *Gammaproteobacteria* dominate dark carbon fixation in coastal sediments. *ISME J.* 10, 1939–1953. doi: 10.1038/ismej.2015.257

Faure, G., Makarova, K. S., and Koonin, E. V. (2019). CRISPR-Cas: complex functional networks and multiple roles beyond adaptive immunity. *J. Mol. Biol.* 431, 3–20. doi: 10.1016/j.jmb.2018.08.030

Gao, S. M., Schippers, A., Chen, N., Yuan, Y., Zhang, M. M., Li, Q., et al. (2020). Depth-related variability in viral communities in highly stratified sulfidic mine tailings. *Microbiome* 8:89. doi: 10.1186/s40168-020-00848-3

Gómez-Santos, N., Glatter, T., Koebnik, R., Świątek-Połatyńska, M. A., and Søgaard-Andersen, L. (2019). A TonB-dependent transporter is required for secretion of protease PopC across the bacterial outer membrane. *Nat. Commun.* 10:1360. doi: 10.1038/s41467-019-09366-9

Hallberg, K. B., Hedrich, S., and Johnson, D. B. (2011). *Acidiferrobacter thiooxydans*, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron-and sulfur-oxidizer of the family Ectothiorhodospiraceae. *Extremophiles* 15, 271–279.

Harms, A., Brodersen, D. E., Mitarai, N., and Gerdes, K. (2018). Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol. Cell* 70, 768–784. doi: 10.1016/j.molcel.2018.01.003

Harrison, A. P. (1982). Genomic and physiological diversity amongst strains of *Thiobacillus ferrooxidans*, and genomic comparison with *Thiobacillus thiooxidans*. *Arch. Microbiol.* 131, 68–76. doi: 10.1007/BF00451501

Hemme, C. L., Green, S. J., Rishishwar, L., Prakash, O., Pettenato, A., Chakraborty, R., et al. (2016). Lateral gene transfer in a heavy metal-contaminated-groundwater microbial community. *mBio* 7:e02234-15. doi: 10.1128/mBio.02234-15

Hernandez-Maldonado, J., Stoneburner, B., Boren, A., Miller, L., Rosen, M., Oremland, R. S., et al. (2016). Genome sequence of the Photoarsenotrophic bacterium *Ectothiorhodospira* sp. Strain BSL-9, isolated from a hypersaline alkaline arsenic-rich extreme environment. *Genome Announc.* 4:e01139-16. doi: 10.1128/genomeA.01139-16

Hsiao, W. W. L., Ung, K., Aeschliman, D., Bryan, J., Finlay, B. B., and Brinkman, F. S. L. (2005). Evidence of a large novel gene pool associated with prokaryotic genomic Islands. *PLoS Genet.* 1:e62. doi: 10.1371/journal.pgen.0010062

Huber, H., and Stetter, K. O. (1989). *Thiobacillus prosperus* sp. nov., represents a new group of halotolerant metal-mobilizing bacteria isolated from a marine geothermal field. *Arch. Microbiol.* 151, 479–485.

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

Issotta, F., Moya-Beltrán, A., Mena, C., Covarrubias, P. C., Thyssen, C., Bellenberg, S., et al. (2018). Insights into the biology of acidophilic members of the *Acidiferrobacteraceae* family derived from comparative genomic analyses. *Res. Microbiol.* 169, 608–617. doi: 10.1016/j.resmic.2018.08.001

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393. doi: 10.1111/j.1574-6976.2008.00136.x

Kazazian, H. H. (2004). Mobile elements: drivers of genome evolution. *Science* 303:1626. doi: 10.1126/science.1089670

Kojima, H., Shinohara, A., and Fukui, M. (2015). *Sulfurifustis variabilis* gen. nov., sp. nov., a sulfur oxidizer isolated from a lake, and proposal of Acidiferrobacteraceae fam. nov. and Acidiferrobacterales ord. nov. *Int. J. Syst. Evol. Microbiol.* 65, 3709–3713. doi: 10.1099/ijsem.0.000479

Kojima, H., Watanabe, T., and Fukui, M. (2016). *Sulfuricaulis limicola* gen. nov., sp. nov., a sulfur oxidizer isolated from a lake. *Int. J. Syst. Evol. Microbiol.* 66, 266–270. doi: 10.1099/ijsem.0.000709

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61. doi: 10.1186/gb-2007-8-4-r61

Langille, M. G. I., Hsiao, W. W. L., and Brinkman, F. S. L. (2008). Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 9:329. doi: 10.1186/1471-2105-9-329

Leal, N. C., Campos, T. L., Rezende, A. M., Docena, C., Mendes-Marques, C. L., de Sá Cavalcanti, F. L., et al. (2020). Comparative genomics of *Acinetobacter baumannii* clinical strains from Brazil reveals polyclonal dissemination and selective exchange of mobile genetic elements associated with resistance genes. *Front. Microbiol.* 11:1176. doi: 10.3389/fmicb.2020.01176

Lee, I., Ouk Kim, Y., Park, S.-C., and Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103. doi: 10.1099/ijsem.0.000760

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018). Virus taxonomy: the database of the international committee on taxonomy of viruses (ICTV). *Nucleic Acids Res.* 46, D708–D717. doi: 10.1093/nar/gkx932

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, 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

Li, L., Liu, Z., Meng, D., Liu, X., Li, X., Zhang, M., et al. (2019). Comparative genomic analysis reveals the distribution, organization, and evolution of metal resistance genes in the genus *Acidithiobacillus*. *Appl. Environ. Microbiol.* 85:e02153-18. doi: 10.1128/AEM.02153-18

Liljeqvist, M., Valdes, J., Holmes, D. S., and Dopson, M. (2011). Draft genome of the psychrotolerant acidophile *Acidithiobacillus ferrivorans* SS3. *J. Bacteriol.* 193, 4304–4305. doi: 10.1128/jb.05373-11

Locher, K. P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct. Mol. Biol.* 23, 487–493. doi: 10.1038/nsmb.3216

Ma, L., Huang, S., Wu, P., Xiong, J., Wang, H., Liao, H., et al. (2021). The interaction of acidophiles driving community functional responses to the re-inoculated chalcopyrite bioleaching process. *Sci. Total Environ.* 798, 149186. doi: 10.1016/j.scitotenv.2021.149186

Ma, L., Wang, H., Wu, J., Wang, Y., Zhang, D., and Liu, X. (2019). Metatranscriptomics reveals microbial adaptation and resistance to extreme environment coupling with bioleaching performance. *Bioresour. Technol.* 280, 9–17. doi: 10.1016/j.biortech.2019.01.117

Maistrenko, O. M., Mende, D. R., Luetge, M., Hildebrand, F., Schmidt, T. S. B., Li, S. S., et al. (2020). Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* 14, 1247–1259. doi: 10.1038/s41396-020-0600-z

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

Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288, 911–940. doi: 10.1006/jmbi.1999.2700

Menikpurage, I. P., Barraza, D., Meléndez, A. B., Strebe, S., and Mera, P. E. (2019). The B12 receptor BtuB alters the membrane integrity of *Caulobacter crescentus*. *Microbiology* 165, 311–323. doi: 10.1099/mic.0.000753

Mitchell, D., Harneit, K., Meyer, G., Sand, W., and Stackebrandt, E. (2004). "Systematic analysis of our culture collection for "genospecies" of *Acidithiobacillus ferrooxidans*, *Acidithiobacillus thiooxidans* and *Leptospirillum ferrooxidans*," in *Biohydrometallurgy: Fundamentals, Technology and Sustainable Development*, eds V. S. T. Ciminelli, O. Garcia Jr., M. C. Teixeira, R. P. de Carvalho, and P. F. Pimentel (Amsterdam: Elsevier), 1369–1378.

Muyzer, G., Sorokin, D. Y., Mavromatis, K., Lapidus, A., Clum, A., Ivanova, N., et al. (2011). Complete genome sequence of "*Thioalkalivibrio sulfidophilus*" HL-EbGr7. *Stand. Genomic Sci.* 4, 23–35. doi: 10.4056/sigs.1483693

Noinaj, N., Guillier, M., Barnard, T. J., and Buchanan, S. K. (2010). TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* 64, 43–60. doi: 10.1146/annurev.micro.112408.134247

Ochsner, A. M., Hemmerle, L., Vonderach, T., Nüssli, R., Bortfeld-Miller, M., Hattendorf, B., et al. (2019). Use of rare-earth elements in the phyllosphere colonizer Methylobacterium extorquens PA1. *Mol. Microbiol.* 111, 1152–1166. doi: 10.1111/mmi.14208

Otsuka, Y. (2016). Prokaryotic toxin–antitoxin systems: novel regulations of the toxins. *Curr. Genet.* 62, 379–382. doi: 10.1007/s00294-015-0557-z

Pablo Cárdenas, J., Ortiz, R., Norris, P. R., Watkin, E., and Holmes, D. S. (2015). Reclassification of '*Thiobacillus prosperus*' Huber and Stetter, 1989 as *Acidihalobacter prosperus* gen. nov., sp. nov., a member of the family Ectothiorhodospiraceae. *Int. J. Syst. Evol. Microbiol.* 65, 3641–3644. doi: 10.1099/ijsem.0.000468

Parte, A. C. (2018). LPSN – List of Prokaryotic names with standing in nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* 68, 1825–1829. doi: 10.1099/ijsem.0.002786

Ramisetty, B. C. M., and Sudhakari, P. A. (2019). Bacterial 'Grounded' prophages: hotspots for genetic renovation and innovation. *Front. Genet.* 10:65. doi: 10.3389/fgene.2019.00065

Rao, R. T., Sharma, S., Sivakumar, N., and Jayakumar, K. (2020). Genomic islands and the evolution of livestock-associated *Staphylococcus aureus* genomes. *Biosci. Rep.* 40:BSR20202287. doi: 10.1042/bsr20202287

Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). Bypassing cultivation to identify bacterial species. *Microbe* 9, 111–118.

Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891. doi: 10.1111/1574-6976.12067

Siguier, P., Perochon, J., Mahillon, J., Lestrade, L., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36.

Song, W., Sun, H.-X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., et al. (2019). Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* 47, W74–W80. doi: 10.1093/nar/gkz380

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

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tapia, P., Flores, F. M., Covarrubias, P. C., Acuña, L. G., Holmes, D. S., and Quatrini, R. (2012). Complete genome sequence of temperate bacteriophage AcaML1 from the extreme acidophile *Acidithiobacillus caldus* ATCC 51756. *J. Virol.* 86, 12452–12453. doi: 10.1128/JVI.02261-12

Thouin, H., Battaglia-Brunet, F., Norini, M.-P., Joulian, C., Hellal, J., Le Forestier, L., et al. (2019). Microbial community response to environmental changes in a technosol historically contaminated by the burning of chemical ammunitions. *Sci. Total Environ.* 697:134108. doi: 10.1016/j.scitotenv.2019.134108

Umezawa, K., Watanabe, T., Miura, A., Kojima, H., and Fukui, M. (2016). The complete genome sequences of sulfur-oxidizing Gamma*proteobacteria Sulfurifustis variabilis* skN76T and Sulfuricaulis limicola HA5T. *Stand. Genom. Sci.* 11:71. doi: 10.1186/s40793-016-0196-0

Unterholzner, S. J., Poppenberger, B., and Rozhon, W. (2013). Toxin–antitoxin systems. *Mobile Genet. Elements* 3:e26219. doi: 10.4161/mge.26219

Valdes, J., Ossandon, F., Quatrini, R., Dopson, M., and Holmes, D. S. (2011). Draft genome sequence of the extremely acidophilic biomining bacterium *Acidithiobacillus thiooxidans* ATCC 19377 provides insights into the evolution of the *Acidithiobacillus* genus. *J Bacteriol* 193, 7003–7004. doi: 10.1128/jb.06281-11

Valdés, J., Pedroso, I., Quatrini, R., Dodson, R. J., Tettelin, H., Blake, R. II, et al. (2008). *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics* 9:597. doi: 10.1186/1471-2164-9-597

Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* 43, 709–730. doi: 10.1080/1040841x.2017.1303661

Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., et al. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142. doi: 10.1186/1471-2105-7-142

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wu, S., You, F., Hall, M., and Huang, L. (2021). Native plant Maireana brevifolia drives prokaryotic microbial community development in alkaline Fe ore tailings under semi-arid climatic conditions. *Sci. Total Environ.* 760, 144019. doi: 10.1016/j.scitotenv.2020.144019

Zhang, X., Liu, X., Liang, Y., Fan, F., Zhang, X., and Yin, H. (2016). Metabolic diversity and adaptive mechanisms of iron- and/or sulfur-oxidizing autotrophic acidophiles in extremely acidic environments. *Environ. Microbiol. Rep.* 8, 738–751. doi: 10.1111/1758-2229.12435

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 February 2022
doi: 10.3389/fmicb.2022.815044

Genomic Insights on Variation Underlying Capsule Expression in Meningococcal Carriage Isolates From University Students, United States, 2015–2016

*Melissa J. Whaley[1], Jeni T. Vuong[1], Nadav Topaz[2], How-Yi Chang[3], Jennifer Dolan Thomas[1], Laurel T. Jenkins[3], Fang Hu[3], Susanna Schmink[1], Evelene Steward-Clark[1], Marsenia Mathis[3], Lorraine D. Rodriguez-Rivera[3], Adam C. Retchless[1], Sandeep J. Joseph[3], Alexander Chen[1], Anna M. Acosta[1], Lucy McNamara[1], Heidi M. Soeters[1], Sarah Mbaeyi[1], Henju Marjuki[1] and Xin Wang[1]**

[1] Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States, [2] CDC Foundation Field Employee assigned to the Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States, [3] IHRC Inc., Contractor to Meningitis and Vaccine Preventable Diseases Branch, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States

OPEN ACCESS

Edited by:
Baolei Jia,
Chung-Ang University, South Korea

Reviewed by:
Paul Balmer,
Pfizer, United States
Peter T. Beernink,
University of California,
San Francisco, United States
Luke R. Green,
The University of Sheffield,
United Kingdom

***Correspondence:**
Xin Wang
gqe8@cdc.gov

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

Received: 14 November 2021
Accepted: 11 January 2022
Published: 17 February 2022

Citation:
Whaley MJ, Vuong JT, Topaz N, Chang HY, Thomas JD, Jenkins LT, Hu F, Schmink S, Steward-Clark E, Mathis M, Rodriguez-Rivera LD, Retchless AC, Joseph SJ, Chen A, Acosta AM, McNamara L, Soeters HM, Mbaeyi S, Marjuki H and Wang X (2022) Genomic Insights on Variation Underlying Capsule Expression in Meningococcal Carriage Isolates From University Students, United States, 2015–2016. Front. Microbiol. 13:815044. doi: 10.3389/fmicb.2022.815044

In January and February 2015, *Neisseria meningitidis* serogroup B (NmB) outbreaks occurred at two universities in the United States, and mass vaccination campaigns using MenB vaccines were initiated as part of a public health response. Meningococcal carriage evaluations were conducted concurrently with vaccination campaigns at these two universities and at a third university, where no NmB outbreak occurred. Meningococcal isolates (N = 1,514) obtained from these evaluations were characterized for capsule biosynthesis by whole-genome sequencing (WGS). Functional capsule polysaccharide synthesis (*cps*) loci belonging to one of seven capsule genogroups (B, C, E, W, X, Y, and Z) were identified in 122 isolates (8.1%). Approximately half [732 (48.4%)] of isolates could not be genogrouped because of the lack of any serogroup-specific genes. The remaining 660 isolates (43.5%) contained serogroup-specific genes for genogroup B, C, E, W, X, Y, or Z, but had mutations in the *cps* loci. Identified mutations included frameshift or point mutations resulting in premature stop codons, missing or fragmented genes, or disruptions due to insertion elements. Despite these mutations, 49/660 isolates expressed capsule as observed with slide agglutination, whereas 45/122 isolates with functional *cps* loci did not express capsule. Neither the variable capsule expression nor the genetic variation in the *cps* locus was limited to a certain clonal complex, except for capsule null isolates (predominantly clonal complex 198). Most of the meningococcal carriage isolates collected from student populations at three US universities were non-groupable as a result of either being capsule null or containing mutations within the capsule locus. Several mutations inhibiting expression of the genes involved with the synthesis and transport of the capsule may be reversible, allowing the bacteria to switch between an encapsulated and non-encapsulated state.

These findings are particularly important as carriage is an important component of the transmission cycle of the pathogen, and understanding the impact of genetic variations on the synthesis of capsule, a meningococcal vaccine target and an important virulence factor, may ultimately inform strategies for control and prevention of disease caused by this pathogen.

Keywords: ***Neisseria meningitidis*****, carriage, capsule, whole genome sequencing, mutations, variation**

INTRODUCTION

Neisseria meningitidis (Nm, meningococcus) is a Gram-negative bacterium that colonizes the human upper respiratory tract. Approximately 10–15% of the general population, and 35% or more of individuals living in close communities (e.g., military or university residences) carry this bacterium without experiencing any symptoms (Stephens, 1999; Oldfield et al., 2017; Soeters et al., 2017). Meningococcal carriage may provide some protection from disease as colonization elicits a mucosal antibody response (Yazdankhah and Caugant, 2004). Meningococcal transmission occurs by direct contact with respiratory secretions from carriers. While mechanisms for the progression from meningococcal carriage to disease are still not well understood, meningococci can breach the mucosal barrier and invade the bloodstream to cause invasive disease such as bacteremia, septicemia, or meningitis, resulting in death in approximately 10–15% of cases (Tzeng and Stephens, 2000; Brandtzaeg and van Deuren, 2012).

As the capsule is a major virulence factor, invasive disease is nearly always caused by meningococci expressing the capsule. Nm is divided into 12 serogroups based on structure of the capsular polysaccharide expressed. Of those, serogroups A, B, C, W, X, and Y cause the majority of invasive disease (Tzeng et al., 2016). Several vaccines are currently available to protect against disease caused by Nm, including quadrivalent meningococcal conjugate vaccines, providing protection against serogroups A, C, W, and Y, and protein-based serogroup B meningococcal (MenB) vaccines, providing protection against serogroup B strains. Non-groupable meningococci (NmNG), or meningococci that do not express capsule, are more commonly recovered from asymptomatic carriers. However, rare cases of invasive disease caused by NmNG have been reported among patients who are immunocompromised, particularly those with complement component deficiency (Ganesh et al., 2017; McNamara et al., 2019).

The meningococcal genes involved in capsule biosynthesis and transport are well characterized and mapped to a region on the chromosome called the capsule polysaccharide synthesis (*cps*) locus (Dolan-Livengood et al., 2003; Harrison et al., 2013). Region A is responsible for capsule polysaccharide synthesis, and genes in this region have been used to categorize isolates into genogroups by either real-time polymerase chain reaction (rt-PCR) or whole genome sequencing (WGS) (Mothershed et al., 2004; World Health Organization, 2011; Marjuki et al., 2019). Regions A, B, and C are required for expression of the serogroup-specific capsule (Harrison et al., 2013). The reduction or loss of meningococcal capsule expression, as a result of phase variation, loss of the entire *cps* operon (capsule null, *cnl*), and/or incorporation of insertion sequences elements (ISEs) within the *cps* locus, may prevent recognition by the host adaptive immune response and potentially result in polysaccharide-based vaccine escape (Dolan-Livengood et al., 2003; Tzeng et al., 2016).

In 2015, meningococcal serogroup B outbreaks occurred at two universities in Oregon (OR) and Rhode Island (RI) in the United States (US). Mass campaigns with a meningococcal serogroup B vaccine were implemented to help control the outbreaks. In conjunction with these campaigns, cross-sectional meningococcal carriage evaluations were conducted at both universities as well as at a third university in RI, where no Nm cases were reported. Most of the carriage isolates collected were non-groupable by rt-PCR and slide agglutination serogrouping (SASG) (McNamara et al., 2017; Soeters et al., 2017; Breakwell et al., 2018). We further characterized these carriage isolates *via* WGS and analyzed the *cps* loci to better understand the possible mechanisms of non-groupability, the potential reversibility of capsule expression, and the likelihood of certain carried isolates causing disease.

MATERIALS AND METHODS

Carriage Isolate Collection

From 10 cross-sectional meningococcal carriage evaluation rounds conducted between February 2015 and May 2016 at the three universities in OR and RI, 8,905 oropharyngeal swabs were collected. A total of 1,514 Nm isolates were obtained from 1,301 unique individuals. Carriage rates were stable between evaluation rounds, ranging between 11 and 24% (McNamara et al., 2017; Soeters et al., 2017; Breakwell et al., 2018). As carriage rates overall and specific to genogroup B were similar among the universities (McNamara et al., 2017; Breakwell et al., 2018), data were aggregated. Carriage evaluation study design and methods have been previously described (McNamara et al., 2017; Soeters et al., 2017; Breakwell et al., 2018). Single colony per participant was selected and subcultured for long-term storage (McNamara et al., 2017; Soeters et al., 2017; Breakwell et al., 2018). All isolates were identified as Nm by Gram stain (BD BBL), oxidase test (Hardy Diagnostics, Santa Maria, CA, United States), *sodC* rt-PCR (Dolan Thomas et al., 2011), and API NH tests (bioMérieux, Durham, NC, United States) (McNamara et al., 2017; Soeters et al., 2017; Breakwell et al., 2018). Isolates were plated once from frozen stocks for further characterization. This evaluation was determined to be public health research, rather than human subjects research; as such, review by the Centers for Disease Control and Prevention (CDC) Institutional Review Board was not required.

Characterization of *Neisseria meningitidis* Capsule Polysaccharide Synthesis Genes by Whole Genome Sequencing

For sequencing on Illumina platforms (HiSeq2500 or MiSeq; San Diego, CA, United States), Nm genomic DNA was prepared with the 5 Prime ArchivePure DNA Purification kit (Gaithersburg, MD, United States). Genomic DNA was further prepared for Illumina sequencing, as previously described, through mechanical shearing and with the dual-index NEBNext Ultra DNA library preparation kits (New England Biolabs Inc., Ipswich, MA, United States) and AMPure XP kit (Beckman Coulter Inc., Indianapolis, IN, United States) (Retchless et al., 2016). Each isolate's final library was loaded on 250-bp pair-end sequencing kits. Illumina reads were trimmed with cutadapt (Martin, 2011) and assembled with SPAdes 3.7.0 (Bankevich et al., 2012). Multilocus sequence typing (MLST) was performed from genome assemblies as previously described (Retchless et al., 2016) to provide sequence type (ST) and clonal complex (CC). The sequence reads for all isolates used in this analysis are available under NCBI BioProject PRJNA533315.

Capsule genes were identified in each assembly as previously described (Marjuki et al., 2019), using allele sequences from the PubMLST *Neisseria* database and Insertion-Sequence (IS) element sequences from the ISFinder database (Altschul et al., 1997; Siguier et al., 2006; Jolley et al., 2018). Capsule polysaccharide biosynthesis and transport genes were identified for each serogroup, specifically in regions A, B, C, and E of the *cps* locus (Harrison et al., 2013). The presence of at least one serogroup-specific gene allowed for genogroup assignment. Isolates lacking any serogroup-specific gene were referred to as genogroup non-groupable. Genogroup non-groupable isolates were further classified by the presence of some *cps* genes (NG_undetermined) or absence of the full *cps* locus (NG_*cnl*). The *cps* locus was examined for completeness and for genetic mutations, such as frameshift or point mutations resulting in premature stop codons (referred to as internal stops), missing or fragmented genes, or disruptions due to insertion elements (Marjuki et al., 2019). For this analysis, phase variation was identified from allele sequences in their phase variable off form as described by PubMLST; these allele sequences contain frameshift mutations arising from length variation from mononucleotide repeats in the sequence. Isolates that did not contain any mutations in their *cps* loci and had all expected capsule genes for the serogroup were designed as containing a "functional *cps* locus" indicating that they are predicted to express the capsule.

Phenotypic Characterization of *Neisseria meningitidis* Capsule

Capsule type and expression in Nm carriage isolates were determined by SASG (World Health Organization, 2011; Breakwell et al., 2018). Isolates were considered groupable if agglutination of 3–4 + intensity was only observed for a single antiserum, except for serogroup Z isolates, which may agglutinate for both E/Z' and Z antisera as described in the product insert (DIFCO, Franklin Lakes, NJ, United States). No agglutination, weak agglutination (1–2 +), polyagglutination (except for serogroup Z isolates), and autoagglutination (as observed in saline) were interpreted as non-groupable. We compared *cps* locus nucleotide sequences from phenotypically non-groupable vs. groupable Nm isolates to assess the association between specific genetic changes and capsule expression. When isolates with identical mutations produced different SASG results or when isolates with functional *cps* loci did not agglutinate, two or more operators performed repeat SASG testing with the same lot of antisera, if available.

RESULTS

Characterization of *Neisseria meningitidis* Capsule by Whole Genome Sequencing and Slide Agglutination Serogrouping

Of the 1,514 meningococcal isolates recovered from the three universities, 782 (51.7%) contained serogroup-specific genes and underwent genogrouping; 732 isolates (48.3%) were NG_*cnl* or NG_undetermined by WGS. A total of 740 isolates (48.9%) had mutations in the *cps* locus, 122 (8.1%) isolates had a functional *cps* locus, and 652 (43%) isolates were capsule null (**Table 1**).

Of the 782 genogroupable isolates, the most common genogroups were E [426/1,514 (28.1%)] and B [251/1,514 (16.6%)]. Genogroups C, W, X, Y, and Z were also detected [4–36 isolates out of 1,514 (0.3–2.4%)]. Genogroup A was not detected in this collection. Most of these isolates (660/782) harbored mutations in the capsule biosynthesis region that may inhibit capsule expression. These mutations included phase variable off, frameshift or point mutations resulting in premature stop

TABLE 1 | Genogrouping of Nm carriage isolates from students at three US universities, 2015–2016.

Genogroup by WGS	No. isolates	No. isolates with functional *cps* loci	No. isolates predicted to not express capsule[a]
B	251	56	195
E	426	43	383
Y	36	15	21
C	31	5	26
W	4	1	3
X	12	1	11
Z	22	1	21
NG_undetermined[b]	80	NA	80
NG_*cnl*	652	NA	652
Total	1,514	122	1,392

cnl, capsule null; cps, capsule polysaccharide synthesis; NA, not applicable; NG, non-groupable; Nm, Neisseria meningitidis; No., number.

[a]*Includes isolates harboring capsule mutations including frameshift, phase variation, point mutations, gene deletions, missing genes, or gene disruptions due to insertion elements, as well as isolates which were capsule null.*

[b]*Isolates containing portions of the cps loci but lacking serogroup-specific genes.*

codons, missing or fragmented genes, and/or disruptions due to insertion elements. The remaining 122 isolates had functional capsule loci for the following genogroups: B [56 (45.9%)], E [43 (35.2%)], Y [15 (12.3%)], C [5 (4.1%)], W [1 (0.8%)], X [1 (0.8%)], and Z [1 (0.8%)] (**Table 1**). By SASG, the majority of isolates [1,388/1,514 (91.7%)] were non-groupable (McNamara et al., 2017; Soeters et al., 2017; Breakwell et al., 2018), and 126 isolates agglutinated with a single antiserum, or two antisera in the case of some serogroup Z isolates. Most of these isolates were either serogroup B [65/126 (51.6%)] or E [53/126 (42.1%)]; serogroups C (n = 1), W (n = 1), Y (n = 5) and Z (n = 1) represented the remaining serogroupable isolates (6.3%).

Genetic Mutations Within the Capsule Polysaccharide Synthesis Loci

The predominant *cps* mutations were assessed among the 740 isolates for each genogroup and the NG_undetermined group (**Supplementary Table 1**). Phase variable off, which commonly presented as frameshifts arising from length variation in simple sequence repeats within the *cps* locus, was found most commonly among genogroup B isolates [84/195 (43.1%)], in comparison to other genogroups [1/26 (3.8%), C; 9/409 (2.2%), E; 0% for all other genogroups]. Deletions of capsule biosynthesis and transport genes were observed among the genogroup C isolates with mutations; 14 of the 26 (53.8%) were missing *cssA*, *cssB*, *cssC*, *ctrA*, *ctrB*, *ctrC*, and *ctrD*. The disruption of a capsule biosynthesis gene by IS*1301* in combination with missing genes or internal stops in other capsule biosynthesis genes was detected in more than a third of genogroup E [146/383 (38.1%)] isolates and most of genogroup Z [19/21 (90.5%)] isolates with mutations. The specific capsule biosynthesis genes disrupted by ISEs for all genogroups are indicated in **Supplementary Table 2**. Internal stops within the capsule biosynthesis genes (*cssC*, *csy*, or *cssA*) were commonly found among genogroup Y isolates [12/21 (57.1%)]; some of these genogroup Y isolates also contained either a fragmented *csy* (2/12) or a *cssA* disrupted by IS*1301* (3/12). All mutations (internal stops or fragmented genes) in the 11 genogroup X isolates affected *csxA*, whereas the three genogroup W isolates with mutations either had an ISE in *cssA* or had multiple biosynthesis genes missing along with an internal stop in the polymerase gene. In addition to lacking serogroup-specific genes, 70/80 (87.5%) genogroup NG_undetermined isolates contained deletions of the biosynthesis genes and portions of the capsule translocation genes in region B.

The Capsule Polysaccharide Synthesis Genes in *Neisseria meningitidis* Isolates With Discrepant Slide Agglutination Serogrouping and Whole Genome Sequencing Results

Isolates with functional *cps* loci are thought to express capsule polysaccharide (Marjuki et al., 2019). A total of 126 isolates expressed capsule polysaccharide as determined by SASG. Among these 126 isolates, 77 were identified to have a functional *cps* locus by WGS [42 (54.5%) for B, 29 (37.6%) for E, 3 (3.9%) for Y, 1 (1.3%) for C, 1 (1.3%) for W, and 1 (1.3%) for Z]. The remaining 49 isolates had various mutations identified in their *cps* loci [23 (46.9%) for B, 24 (49.0%) for E, and 2 (4.1%) for Y]. Internal stops, ISEs, and phase variable off were commonly found among these isolates (**Table 2**). Of the 23 serogroup B isolates, 20 contained *csb* alleles that were determined to be in the phase variable off state based on their genome sequence. Most of the serogroup E isolates contained internal stop codons in either *cseA* or *cseD* genes. One serogroup Y isolate had an internal stop in capsule biosynthesis gene, whereas one had an ISE.

The mutations identified among the 47 serogroupable isolates were also detected among non-serogroupable isolates (**Supplementary Table 3**), indicating that the same mutation may or may not prevent capsule expression as determined by SASG. One of the two (50%) genogroup Y isolates agglutinated with Y antisera when both isolates possessed an internal stop in the same location of the *cssA*. One of 24 (4.2%) genogroup B isolates with an internal stop in the same location of *cssA* expressed a B capsule, as well as 19 of 64 (30%) genogroup B isolates with *csb* alleles presumably in the phase variable off state. Six of 48 (11.1%) genogroup E isolates containing a premature internal stop in *cseA* and 13/16 (81.3%) of genogroup E isolates with premature internal stop in *cseD* agglutinated with the E antisera. Among 122 isolates containing a functional *cps* locus, 45/122 (36.9%) isolates were non-groupable by SASG (**Table 1** and **Supplementary Table 1**), including genogroups B [14 (31.1%)], E [14 (31.1%)], Y [12 (26.7%)], C [4 (8.9%)], and X [1 (2.2%)]. As previously described, the remaining 77 isolates with functional *cps* loci were serogroupable.

TABLE 2 | The *cps* mutations in Nm carriage isolates with discrepant capsule typing results.

Serogroup by SASG (No.)	WGS analysis	No. (%)
B (23)	Internal stop in *csb*	1 (4.3%)
	Internal stop in *cssA*	1 (4.3%)
	Internal stop in *cssA*, phase variable off[a] in *csb*	1 (4.3%)
	Internal stop in *ctrE*	1 (4.3%)
	Phase variable off in *csb*	19 (82.6%)
E (24)	*cseD* disrupted by IS*1301*	3 (12.5%)
	Internal stop in *cseA*	6 (25%)
	Internal stop in *cseD*	10 (41.7%)
	Missing *cseA*, missing *cseB*, missing *cseC*, missing *cseD*	1 (4.1%)
	cseG fragmented (31.89% coverage), *ctrA* fragmented (20.98% coverage), internal stop in *cseG*, missing *cseA*, missing *cseB*, missing *cseC*, missing *cseD*, missing *cseE*	1 (4.1%)
	Internal stop in *cseC*, internal stop in *ctrA*	1 (4.1%)
	Internal stop in *cseD*, internal stop in *cseG*	2 (8.3%)
Y (2)	Internal stop in *cssA*	1 (50%)
	cssC disrupted by IS*1655*	1 (50%)

Nm, Neisseria meningitidis; No., number of isolates; SASG, slide agglutination serogrouping; WGS, whole genome sequencing.
[a]*Defined as frameshifts arising from length variation in mononucleotide repeats.*

Association Between Genogroup and Clonal Complex

The CC distribution by genogroup is shown in **Figure 1**. Most CCs were unique to a genogroup, except for CC1157, CC162, CC167, CC174, CC175, CC178, CC213, CC23, CC269, CC32, CC35, CC44/41, CC60, CC865, and unassigned CC (ST-5953). The most diverse CC by genogroup, CC1157, was the predominant CC for genogroup E isolates [232/426 (54.5%)] and was one of the most common CCs for genogroup X isolates [5/12 (41.7%)]. The other common genogroup X CC was CC865 [5/12 (41.7%)]. The predominant CCs for the other genogroups were as follows: CC41/44 for genogroup B [94/251 (37.5%)], CC269 for genogroup C [15/31 (48.3%)], CC23 for genogroup Y [21/36 (58.3%)], unassigned CC (ST-5953) for genogroup Z [7/22 (31.8%)], CC35 for genogroup NG_undetermined isolates [34/80 (42.5%)], and CC198 for genogroup NG_*cnl* isolates [413/652 (63.3%)]. Genogroup W isolates belonged to either CC175 or CC22.

DISCUSSION

In this study, we characterized the capsule type and expression of Nm isolates obtained from carriage evaluations at three US universities in 2015–2016 and identified genetic variations within the *cps* loci of these isolates. Most of these isolates were genetically non-groupable because of lack of serogroup-specific genes (NG_undetermined) or the whole *cps* locus (NG_*cnl*), followed by genogroups E and B. A majority of these isolates contained genetic mutations such as phase variable off, frameshift or point mutations, missing or fragmented genes, and/or disruptions due to insertion elements within the *cps* loci. A functional *cps* locus was detected only in a very small proportion of Nm carriage isolates.

While genogroup composition among Nm carriage isolates is similar to other carriage studies, E was the most common genogroup among genogroupable isolates, representing 26% of the collection. In contrast, carriage evaluations conducted at universities in Europe and South America identified genogroup B as the most common genogroup; genogroup E isolates were between 2 and 9% of their collection (Gasparini et al., 2014; Rodriguez et al., 2014; Cassio de Moraes et al., 2015; Jeppesen et al., 2015; Díaz et al., 2016; Tryfinopoulou et al., 2016; Moura et al., 2017; van Ravenhorst et al., 2017; Oldfield et al., 2018). In earlier US carriage evaluations, genogroup E carriage was 10% among high school students (Harrison et al., 2015). The higher proportion of genogroup E isolates may be specific to these US university populations at the time of collection. In a review, meningococcal carriage differed in serogroup distribution within and between European countries, across age groups, and over time (Soriano-Gabarró et al., 2011). Serological methods, such as SASG, were more often used than molecular methods in earlier carriage evaluations (Soriano-Gabarró et al., 2011). The carriage of genogroup E and other rare genogroups may be underestimated in earlier carriage studies as a result of the limited availability of antiserum for these serogroups (Soriano-Gabarró et al., 2011; Harrison et al., 2013; Rodrigues et al., 2015). Genogroup C isolates were rarely detected in carriage evaluations conducted during a time when NmC conjugate vaccines were widely available (Gasparini et al., 2014; Harrison et al., 2015; Moreno et al., 2015; Rodrigues et al., 2015; Tryfinopoulou et al., 2016; Moura et al., 2017; McMillan et al., 2019). Consistently, genogroup C represents only a small proportion of carriage isolates in our collection, which may be due to the routine recommendations for use of meningococcal ACWY vaccines among adolescents (Mbaeyi et al., 2020).

Predominant *cps* mutations for each genogroup were assessed to understand their impact on capsule expression. While some genetic mutations result in irreversible loss of capsule expression, some mutations, such as frameshifts and point mutations that were commonly found in the capsule biosynthesis genes of genogroup B and E isolates in our collection, may serve as an on–off switch to regulate capsule expression under different conditions. This on–off switch can down-regulate capsule expression to allow Nm colonization during carriage and up-regulate it to allow progression to disease (Hammerschmidt et al., 1996; Weber et al., 2006; Tzeng et al., 2016) by introducing or removing premature internal stops. In addition, we also observed that several genogroup B isolates with *csb* in the phase variable off state still expressed a B capsule, suggesting phase variable status may change during isolation and culturing of Nm (Lucidarme et al., 2013; Siena et al., 2016). Capsule expression has also been detected by phenotypic methods for isolates with ISEs in *cssA*, *cssE*, *cssF*, or *csb* in earlier studies (Hammerschmidt et al., 1996; Weber et al., 2006; Germinario et al., 2010; Kugelberg et al., 2010; Loh et al., 2013; Jones et al., 2016; Tzeng et al., 2016). In this study, agglutination was observed among genogroup E isolates containing internal stops in *cseD*, suggesting that the disruption *cseD* did not prevent capsule expression. Another explanation may be that the function of *cseD*, lipopolysaccharide–KDO synthesis, is complemented elsewhere in the genome (Harrison et al., 2013; Tzeng et al., 2016). An unlinked gene complementation may also explain agglutination for genogroup B and E isolates with internal stops in the biosynthesis genes. One of three genogroup B isolates with an internal stop in *csb* agglutinated, which was also observed for three invasive serogroup B isolates in a previous study (Marjuki et al., 2019). It warrants further investigation whether the different position of the stop codon(s) may possibly have permitted translation of gene products among these isolates, resulting in some capsule expression.

A subset of isolates belonging to genogroups B and Y with functional *cps* loci did not agglutinate with the capsule-specific antisera (non-groupable phenotype). Similar lack of agglutination has been observed in previous studies evaluating both invasive and carriage isolates among university students and general populations (Germinario et al., 2010; Gasparini et al., 2014; Jones et al., 2016; van Ravenhorst et al., 2017; Marjuki et al., 2019). These isolates may have had a biosynthesis gene in the phase variable on state when sequenced and changed to off state when tested for agglutination, as a result of subculturing a single colony (Lucidarme et al., 2013; Siena et al., 2016). To assess the impact of culture preparation on phase variation–mediated capsule expression, we are sequencing

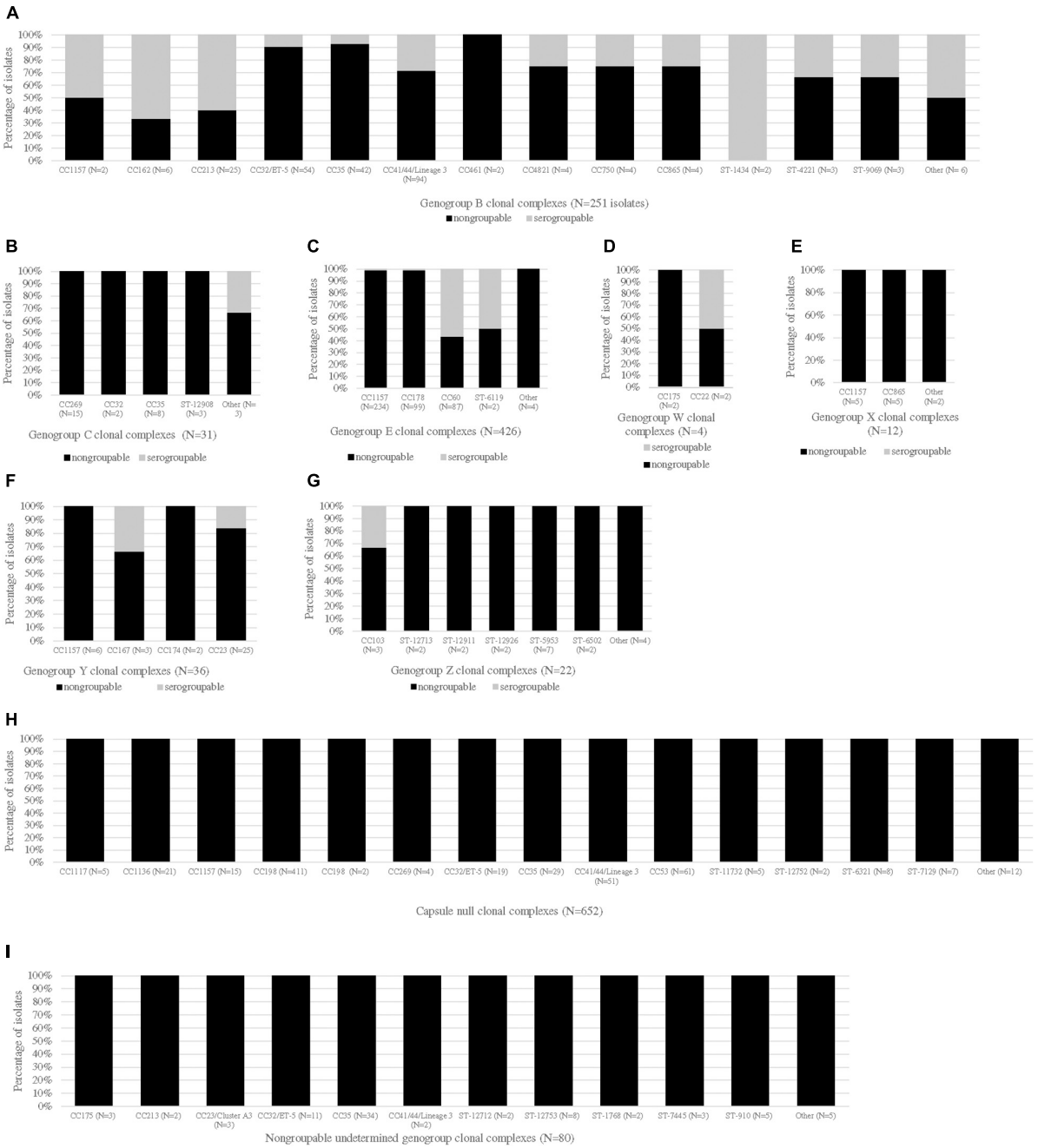


FIGURE 1 | CC distribution of Nm carriage isolates by genogroup and serogroup. The proportions of serogroupable (shown in gray) and non-serogroupable (shown in black) meningococcal isolates were presented for each genogroup-specific clonal complex (CC): genogroup B **(A)**, genogroup C **(B)**, genogroup E **(C)**, genogroup W **(D)**, genogroup X **(E)**, genogroup Y **(F)**, genogroup Z **(G)**, capsule null **(H)**, and non-groupable undetermined genogroup **(I)**. If a CC was unassigned, the sequence type (ST) was provided. The "other" category represents the genogroup-specific CCs that contained only one isolate. The serogroupable isolates belonging to the "other" category were as follows: serogroup B: CC174, ST-11860 and ST-948; and serogroup C: ST-12922.

multiple colonies from a single stock under the same conditions described in the methods (Siena et al., 2016) and testing their capsule expression using agglutination and other methods. We are also examining non-capsule genes to see if they have an impact on capsule expression. These studies in progress may help address the discrepancies we observed. Other reasons for this observation could be that the expression of capsule was at a level so low that was undetectable by SASG (Mothershed et al., 2004; Soriano-Gabarró et al., 2011; Jeppesen et al., 2015; Gilca et al., 2018). SASG testing was repeated on these isolates by different operators and yielded similar results. Although SASG is a common method for serogrouping

Nm with relatively good sensitivity and specificity (van der Ende et al., 1995), we have noted differences between or within lots of antiserum, which unavoidably affect the performance of this test and lead to discrepancies when compared with other test results such as WGS. We have conducted preliminary testing to investigate other phenotypic tests, but so far, discrepancies have not resolved by other methods; this is in agreement with Jones et al. (2016), who assessed specificity and sensitivity of agglutination and other phenotypic tests (flow cytometry, live-cell phenotypic assay, and dot blotting) in comparison to WGS. Others also have observed that, even with the use of other phenotypic assays (Ouchterlony and live-cell phenotypic assay), some genogroup B carriage isolates with functional *cps* loci remained non-groupable (van Ravenhorst et al., 2017; Gilca et al., 2018). Genes adjacent to the *cps* locus and non-coding regions outside of the *cps* locus may influence capsule expression (Swartley et al., 1997). The inactivation of KpsF, encoded by a gene outside of the *cps* locus, was shown to reduce capsule expression (Tzeng et al., 2016). However, the carriage isolates studied here contained an intact *kpsF* (data not shown), but this does not rule out the possibility of the gene expression being inactivated. Further analysis of non-*cps* genes and non-coding regions is warranted to better understand the non-groupability of isolates with functional *cps* loci.

In comparison to other carriage studies (Soriano-Gabarró et al., 2011; Gasparini et al., 2014; Read et al., 2014; Rodriguez et al., 2014; Cassio de Moraes et al., 2015; Jeppesen et al., 2015; Moreno et al., 2015; Díaz et al., 2016; Tryfinopoulou et al., 2016; Moura et al., 2017; Oldfield et al., 2017, 2018; Tekin et al., 2017; van Ravenhorst et al., 2017; Gilca et al., 2018; Terranova et al., 2018; McMillan et al., 2019), similar CC distributions among genogroups were observed. Capsule null isolates commonly belonged to CC198 and rarely belonged to CCs associated with invasive disease (Gasparini et al., 2014; Rodriguez et al., 2014; Cassio de Moraes et al., 2015; Moreno et al., 2015; Rodrigues et al., 2015; Tryfinopoulou et al., 2016; Moura et al., 2017; Oldfield et al., 2018; McMillan et al., 2019). The majority of genogroup Y isolates from our collection belonged to CC23, similar to the US carriage evaluation among high school students (Harrison et al., 2015) and to university carriage studies in Europe (Gasparini et al., 2014; Rodrigues et al., 2015; Oldfield et al., 2018), South America (Moreno et al., 2015; Moura et al., 2017), and Australia (McMillan et al., 2019). Genogroup B isolates predominantly belonged to CC41/44 (Soriano-Gabarró et al., 2011; Rodriguez et al., 2014; Jeppesen et al., 2015; Moreno et al., 2015; Rodrigues et al., 2015; Tryfinopoulou et al., 2016; Tekin et al., 2017; Gilca et al., 2018; McMillan et al., 2019). Genogroup C in this evaluation belonged to CCs commonly associated with *N. meningitidis* serogroup B disease (CC269, CC32, CC35, and CC41/44). Genogroup C isolates belonging to CC41/44 and CC35 were also detected among university carriage isolates collected in Brazil (Cassio de Moraes et al., 2015; Moura et al., 2017) and Colombia (Moreno et al., 2015). CC269 genogroup C isolates appear to be uniquely detected in our collection.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, BioProject PRJNA533315.

AUTHOR CONTRIBUTIONS

MW, JV, NT, HM, and XW conceptualized, performed the analysis, and wrote the manuscript. HC, JT, LJ, FH, SS, ES-C, MM, and LR-R involved in testing of the samples. AA, LM, HS, and SM involved in managing and supervising the carriage studies. AR, SJ, and AC involved in data analysis, interpretation, and data management. XW supervised the project and the wrote of the manuscript. All authors reviewed the manuscript and provided critical feedback.

FUNDING

This work was supported by the Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

We thank our state and local collaborators, the CDC Genome Sequencing Laboratory in the Biotechnology Core Facility Branch, and members of the CDC Meningitis and Vaccine Preventable Diseases Branch. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.815044/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. doi: 10.1093/nar/25.17.3389

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

Brandtzaeg, P., and van Deuren, M. (2012). Classification and pathogenesis of meningococcal infections. *Methods Mol. Biol.* 799, 21–35. doi: 10.1007/978-1-61779-346-2_2

Breakwell, L., Whaley, M., Khan, U. I., Bandy, U., Alexander-Scott, N., Dupont, L., et al. (2018). Meningococcal carriage among a university student population – United States, 2015. *Vaccine* 36, 29–35. doi: 10.1016/j.vaccine.2017.11.040

Cassio de Moraes, J., Kemp, B., de Lemos, A. P., Outeiro Gorla, M. C., Lemes Marques, E. G., Ferreira Mdo, C., et al. (2015). Prevalence, risk factors and molecular characteristics of meningococcal carriage among Brazilian adolescents. *Pediatric Infect. Dis. J.* 34, 1197–1202. doi: 10.1097/INF.0000000000000853

Díaz, J., Cárcamo, M., Seoane, M., Pidal, P., Cavada, G., Puentes, R., et al. (2016). Prevalence of meningococcal carriage in children and adolescents aged 10-19 years in Chile in 2013. *J. Infect. Public Health* 9, 506–515. doi: 10.1016/j.jiph.2015.12.011

Dolan Thomas, J., Hatcher, C. P., Satterfield, D. A., Theodore, M. J., Bach, M. C., Linscott, K. B., et al. (2011). sodC-based real-time PCR for detection of *Neisseria meningitidis*. *PLoS One* 6:e19361. doi: 10.1371/journal.pone.0019361

Dolan-Livengood, J. M., Miller, Y. K., Martin, L. E., Urwin, R., and Stephens, D. S. (2003). Genetic basis for nongroupable *Neisseria meningitidis*. *J. Infect. Dis.* 187, 1616–1628. doi: 10.1086/374740

Ganesh, K., Allam, M., Wolter, N., Bratcher, H. B., Harrison, O. B., Lucidarme, J., et al. (2017). Molecular characterization of invasive capsule null *Neisseria meningitidis* in South Africa. *BMC Microbiol.* 17:40. doi: 10.1186/s12866-017-0942-5

Gasparini, R., Comanducci, M., Amicizia, D., Ansaldi, F., Canepa, P., Orsi, A., et al. (2014). Molecular and serological diversity of *Neisseria Meningitidis* carrier strains isolated from Italian students aged 14 to 22 years. *J. Clin. Microbiol.* 52, 1901–1910. doi: 10.1128/JCM.03584-13

Germinario, C., Tafuri, S., Napoli, C., Montagna, M. T., Balducci, M. T., Fortunato, F., et al. (2010). Young-adult carriers of *Neisseria meningitidis* in Puglia (Italy): will the pattern of circulating meningococci change following the introduction of meningococcal serogroup C conjugate vaccines? *Hum. Vaccin.* 6, 1025–1027. doi: 10.4161/hv.6.12.13145

Gilca, R., De Wals, P., Nolan, S. M., Kitchin, N., Eiden, J. J., Jiang, Q., et al. (2018). A longitudinal epidemiology study of meningococcal carriage in students 13 to 25 years old in Quebec. *mSphere* 3:e00427-18. doi: 10.1128/mSphere.00427-18

Hammerschmidt, S., Müller, A., Sillmann, H., Miihlenhoff, M., Borrow, R., Fox, A., et al. (1996). Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (siaD): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol. Microbiol.* 20, 1211–1220. doi: 10.1111/j.1365-2958.1996.tb02641.x

Harrison, L. H., Shutt, K. A., Arnold, K. E., Stern, E. J., Pondo, T., Kiehlbauch, J. A., et al. (2015). Meningococcal carriage among georgia and maryland high school students. *J. Infect. Dis.* 211, 1761–1768. doi: 10.1093/infdis/jiu679

Harrison, O. B., Claus, H., Jiang, Y., Bennett, J. S., Bratcher, H. B., Jolley, K. A., et al. (2013). Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerg. Infect. Dis.* 19, 566–573. doi: 10.3201/eid1904.111799

Jeppesen, C. A., Snape, M. D., Robinson, H., Gossger, N., John, T. M., Voysey, M., et al. (2015). Meningococcal carriage in adolescents in the United Kingdom to inform timing of an adolescent vaccination strategy. *J. Infect.* 71, 43–52. doi: 10.1016/j.jinf.2015.02.006

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3, 124–124. doi: 10.12688/wellcomeopenres.14826.1

Jones, C. H., Mohamed, N., Rojas, E., Andrew, L., Hoyos, J., Hawkins, J. C., et al. (2016). Comparison of phenotypic and genotypic approaches to capsule typing of *Neisseria meningitidis* by use of invasive and carriage isolate collections. *J. Clin. Microbiol.* 54, 25–34. doi: 10.1128/JCM.01447-15

Kugelberg, E., Gollan, B., Farrance, C., Bratcher, H., Lucidarme, J., Ibarz-Pavón, A. B., et al. (2010). The influence of IS1301 in the capsule biosynthesis locus on meningococcal carriage and disease. *PLoS One* 5:e9413. doi: 10.1371/journal.pone.0009413

Loh, E., Kugelberg, E., Tracy, A., Zhang, Q., Gollan, B., Ewles, H., et al. (2013). Temperature triggers immune evasion by *Neisseria meningitidis*. *Nature* 502, 237–240. doi: 10.1038/nature12616

Lucidarme, J., Findlow, J., Chan, H., Feavers, I. M., Gray, S. J., Kaczmarski, E. B., et al. (2013). The distribution and '*in vivo*' phase variation status of haemoglobin receptors in invasive meningococcal serogroup b disease: genotypic and phenotypic analysis. *PLoS One* 8:e76932. doi: 10.1371/journal.pone.0076932

Marjuki, H., Topaz, N., Rodriguez-Rivera, L. D., Ramos, E., Potts, C. C., Chen, A., et al. (2019). Whole-genome sequencing for characterization of capsule locus and prediction of serogroup of invasive meningococcal isolates. *J. Clin. Microbiol.* 57:e01609-18. doi: 10.1128/JCM.01609-18

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.1089/cmb.2017.0096

Mbaeyi, S. A., Bozio, C. H., Duffy, J., Rubin, L. G., Hariri, S., Stephens, D. S., et al. (2020). Meningococcal vaccination: recommendations of the advisory committee on immunization practices, United States, 2020. *MMWR Recomm. Rep.* 69, 1–41. doi: 10.15585/mmwr.rr6909a1

McMillan, M., Walters, L., Mark, T., Lawrence, A., Leong, L. E. X., Sullivan, T., et al. (2019). B Part of It study: a longitudinal study to assess carriage of *Neisseria meningitidis* in first year university students in South Australia. *Hum. Vaccin. Immunother.* 15, 987–994. doi: 10.1080/21645515.2018.1551672

McNamara, L. A., Potts, C. C., Blain, A., Topaz, N., Apostol, M., Alden, N. B., et al. (2019). Invasive meningococcal disease due to nongroupable *Neisseria meningitidis*-active bacterial core surveillance sites, 2011-2016. *Open Forum Infect. Dis.* 6:ofz190. doi: 10.1093/ofid/ofz190

McNamara, L. A., Thomas, J. D., MacNeil, J., Chang, H. Y., Day, M., Fisher, E., et al. (2017). Meningococcal carriage following a vaccination campaign with MenB-4C and MenB-FHbp in response to a university serogroup B meningococcal disease outbreak—Oregon, 2015–2016. *J. Infect. Dis.* 216, 1130–1140. doi: 10.1093/infdis/jix446

Moreno, J., Hidalgo, M., Duarte, C., Sanabria, O., Gabastou, J. M., and Ibarz-Pavon, A. B. (2015). Characterization of carriage isolates of *Neisseria meningitides* in the adolescents and young adults population of Bogota (Colombia). *PLoS One* 10:e0135497. doi: 10.1371/journal.pone.0135497

Mothershed, E. A., Sacchi, C. T., Whitney, A. M., Barnett, G. A., Ajello, G. W., Schmink, S., et al. (2004). Use of real-time PCR to resolve slide agglutination discrepancies in serogroup identification of *Neisseria meningitidis*. *J. Clin. Microbiol.* 42, 320–328. doi: 10.1128/JCM.42.1.320-328.2004

Moura, A. R. S. S., Kretz, C. B., Ferreira, I. E., Nunes, A. M. P. B., de Moraes, J. C., Reis, M. G., et al. (2017). Molecular characterization of *Neisseria meningitidis* isolates recovered from 11-19-year-old meningococcal carriers in Salvador, Brazil. *PLoS One* 12:e0185038. doi: 10.1371/journal.pone.0185038

Oldfield, N. J., Cayrou, C., AlJannat, M. A. K., Al-Rubaiawi, A. A. A., Green, L. R., Dada, S., et al. (2017). Rise in group W meningococcal carriage in university students, United Kingdom. *Emerg. Infect. Dis.* 23, 1009–1011. doi: 10.3201/eid2306.161768

Oldfield, N. J., Green, L. R., Parkhill, J., Bayliss, C. D., and Turner, D. P. J. (2018). Limited impact of adolescent meningococcal ACWY vaccination on *Neisseria meningitidis* serogroup W carriage in university students. *J. Infect. Dis.* 217, 608–616. doi: 10.1093/infdis/jix596

Read, R. C., Baxter, D., Chadwick, D. R., Faust, S. N., Finn, A., Gordon, S. B., et al. (2014). Effect of a quadrivalent meningococcal ACWY glycoconjugate or a serogroup B meningococcal vaccine on meningococcal carriage: an observer-blind, phase 3 randomised clinical trial. *Lancet* 384, 2123–2131. doi: 10.1016/S0140-6736(14)60842-4

Retchless, A. C., Hu, F., Ouédraogo, A.-S., Diarra, S., Knipe, K., Sheth, M., et al. (2016). The establishment and diversification of epidemic-associatedserogroup W meningococcus in the African Meningitis Belt, 1994 to 2012. *mSphere* 1, e201–e216. doi: 10.1128/mSphere.00201-16

Rodrigues, F., Morales-Aza, B., Christensen, H., Giles, J., Ferreira, M., Sikora, P., et al. (2015). Oropharyngeal carriage of meningococcus in portugal by group and clonal complex 6 years after adolescent vaccine campaign. *Pediatric Infect. Dis. J.* 34, 1267–1269. doi: 10.1097/INF.0000000000000860

Rodriguez, P., Alvarez, I., Torres, M. T., Diaz, J., Bertoglia, M. P., Carcamo, M., et al. (2014). Meningococcal carriage prevalence in university students, 1824 years of age in Santiago, Chile. *Vaccine* 32, 5677–5680. doi: 10.1016/j.vaccine.2014.08.015

Siena, E., D'Aurizio, R., Riley, D., Tettelin, H., Guidotti, S., Torricelli, G., et al. (2016). *In-silico* prediction and deep-DNA sequencing validation indicate phase variation in 115 *Neisseria meningitidis* genes. *BMC genomics* 17:843. doi: 10.1186/s12864-016-3185-1

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34(Suppl._1), D32–D36. doi: 10.1093/nar/gkj014

Soeters, H. M., Whaley, M., Alexander-Scott, N., Kanadanian, K. V., MacNeil, J. R., Martin, S. W., et al. (2017). Meningococcal carriage evaluation in response to a serogroup B meningococcal disease outbreak and mass vaccination campaign

at a college—Rhode Island, 2015–2016. *Clin. Infect. Dis.* 64, 1115–1122. doi: 10.1093/cid/cix091

Soriano-Gabarró, M., Wolter, J., Hogea, C., and Vyse, A. (2011). Carriage of *Neisseria meningitidis* in Europe: a review of studies undertaken in the region. *Exp. Rev. Anti Infect. Ther.* 9, 761–774. doi: 10.1586/eri.11.89

Stephens, D. S. (1999). Uncloaking the meningococcus: dynamics of carriage and disease. *Lancet* 353, 941–942. doi: 10.1016/S0140-6736(98)00279-7

Swartley, J. S., Marfin, A. A., Edupuganti, S., Liu, L.-J., Cieslak, P., Perkins, B., et al. (1997). Capsule switching of *Neisseria?meningitidis*. *Proc. Natl. Acad. Sci.* 94, 271–276.

Tekin, R. T., Dinleyici, E. C., Ceyhan, M., Karbuz, A., Salman, N., Sutçu, M., et al. (2017). The prevalence, serogroup distribution and risk factors of meningococcal carriage in adolescents and young adults in Turkey. *Hum. Vaccin. Immunother.* 13, 1182–1189. doi: 10.1080/21645515.2016.1268304

Terranova, L., Principi, N., Bianchini, S., Di Pietro, G., Umbrello, G., Madini, B., et al. (2018). *Neisseria meningitidis* serogroup B carriage by adolescents and young adults living in Milan, Italy: prevalence of strains potentially covered by the presently available meningococcal B vaccines. *Hum. Vaccin. Immunother.* 14, 1070–1074. doi: 10.1080/21645515.2018.1450121

Tryfinopoulou, K., Kesanopoulos, K., Xirogianni, A., Marmaras, N., Papandreou, A., Papaevangelou, V., et al. (2016). Meningococcal carriage in military recruits and university students during the pre MenB vaccination era in Greece (2014-2015). *PLoS One* 11:e0167404. doi: 10.1371/journal.pone.0167404

Tzeng, Y. L., and Stephens, D. S. (2000). Epidemiology and pathogenesis of *Neisseria meningitidis*. *Microbes Infect.* 2, 687–700. doi: 10.1016/s1286-4579(00)00356-7

Tzeng, Y.-L., Thomas, J., and Stephens, D. S. (2016). Regulation of capsule in *Neisseria meningitidis*. *Crit. Rev. Microbiol.* 42, 759–772.

van der Ende, A., Schuurman, I. G., Hopman, C. T., Fijen, C. A., and Dankert, J. (1995). Comparison of commercial diagnostic tests for identification of serogroup antigens of *Neisseria meningitidis*. *J. Clin. Microbiol.* 33, 3326–3327. doi: 10.1128/jcm.33.12.3326-3327.1995

van Ravenhorst, M. B., Bijlsma, M. W., van Houten, M. A., Struben, V. M. D., Anderson, A. S., Eiden, J., et al. (2017). Meningococcal carriage in Dutch adolescents and young adults;a cross-sectional and longitudinal cohort study. *Clin. Microbiol. Infect.* 23, 573.e1–573.e7. doi: 10.1016/j.cmi.2017.02.008

Weber, M. V., Claus, H., Maiden, M. C., Frosch, M., and Vogel, U. (2006). Genetic mechanisms for loss of encapsulation in polysialyltransferase-gene-positive meningococci isolated from healthy carriers. *Int. J. Med. Microbiol.* 296, 475–484. doi: 10.1016/j.ijmm.2006.05.004

World Health Organization (2011). *Laboratory Methods for the Diagnosis of Meningitis Caused by Neisseria meningitidis, Streptococcus pneumoniae, and Haemophilus influenzae, second edn*. Geneva: World Health Organization.

Yazdankhah, S. P., and Caugant, D. A. (2004). *Neisseria meningitidis*: an overview of the carriage state. *J. Med. Microbiol.* 53(Pt 9), 821–832. doi: 10.1099/jmm.0.45529-0

Conflict of Interest: HC, LJ, FH, MM, LR-R, and SJ were employed by the IHRC Inc.

The remaining 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/fmicb.2022.793612

Horizontal Gene Transfer of Fluoroquinolone Resistance-Conferring Genes From Commensal *Neisseria* to *Neisseria gonorrhoeae*: A Global Phylogenetic Analysis of 20,047 Isolates

Sheeba Santhini Manoharan-Basil[1], Natalia González[1], Jolein Gyonne Elise Laumen[1] and Chris Kenyon[1,2]*

[1]Department of Clinical Sciences, Institute of Tropical Medicine Antwerp, Antwerp, Belgium, [2]Department of Medicine, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:
Baolei Jia,
Chung-Ang University, South Korea

Reviewed by:
Oscar Q. Pich,
Parc Taulí Foundation, Spain
Dmitry Gryadunov,
Engelhardt Institute of Molecular Biology (RAS), Russia
Leshan Xiu,
Shanghai Jiao Tong University, China

***Correspondence:**
Sheeba Santhini Manoharan-Basil
sbasil@itg.be

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

Received: 12 October 2021
Accepted: 19 January 2022
Published: 17 March 2022

Citation:
Manoharan-Basil SS, González N, Laumen JGE and Kenyon C (2022) Horizontal Gene Transfer of Fluoroquinolone Resistance-Conferring Genes From Commensal Neisseria to Neisseria gonorrhoeae: A Global Phylogenetic Analysis of 20,047 Isolates. Front. Microbiol. 13:793612. doi: 10.3389/fmicb.2022.793612

Antimicrobial resistance in *Neisseria gonorrhoeae* is an important global health concern. The genetically related commensal *Neisseria* act as a reservoir of resistance genes, and horizontal gene transfer (HGT) has been shown to play an important role in the genesis of resistance to cephalosporins and macrolides in *N. gonorrhoeae*. In this study, we evaluated if there was evidence of HGT in the genes *gyrA/gyrB* and *parC/parE* responsible for fluoroquinolone resistance. Even though the role of *gyrB* and *parE* in quinolone resistance is unclear, the subunits *gyrB* and *parE* were included as zoliflodacin, a promising new drug to treat *N. gonorrhoeae* targets the *gyrB* subunit. We analyzed a collection of 20,047 isolates; 18,800 *N. gonorrhoeae*, 1,238 commensal *Neisseria* spp., and nine *Neisseria meningitidis*. Comparative genomic analyses identified HGT events in genes, *gyrA*, *gyrB*, *parC*, and *parE*. Recombination events were predicted in *N. gonorrhoeae* and *Neisseria* commensals. *Neisseria lactamica*, *Neisseria macacae*, and *Neisseria mucosa* were identified as likely progenitors of the HGT events in *gyrA*, *gyrB*, and *parE*, respectively.

Keywords: HGT, *gyrA*, *gyrB*, *parC*, *parE*, *Neisseria gonorrhoeae*, commensal *Neisseria*

INTRODUCTION

Neisseria gonorrhoeae, a gram-negative diplococcus, is an obligate human pathogen that causes the sexually transmitted infection gonorrhea (Ligon, 2005; Tapsall et al., 2009). *Neisseria gonorrhoeae* has been remarkably adept at developing resistance to antimicrobials, including β-lactams, macrolides, and fluoroquinolones. This has resulted in *N. gonorrhoeae* being included as a high priority pathogen by the WHO (Willcox, 1970; CDC, 1985; Unemo and Shafer, 2011; Chesson et al., 2014; Blair et al., 2015; Tacconelli and Magrini, 2017). Current treatment guidelines typically advocate using only ceftriaxone or in combination with azithromycin (Unemo et al., 2020). Increasing incidence of resistance to both ceftriaxone and azithromycin have led

to novel treatment strategies. A particularly promising novel agent, zoliflodacin, a spiropyrimidinetrione, whose target site is *gyrB* is currently being evaluated in phase III studies following successful phase II studies (Alm et al., 2015; Taylor et al., 2018; Bradford et al., 2020). Recent trials have also established that ciprofloxacin (CIP) can be used to treat gonorrhea if genotyping confirms the absence of *gyrA* mutations (Allan-Blitz et al., 2017). This renewed interest in agents targeting the DNA gyrases led to the current study to assess if *gyrA/B* and *parC/E* mutations in *N. gonorrhoeae* can be acquired by horizontal gene transfer (HGT).

The first-generation quinolone, nalidixic acid, was first used in 1962 (Lesher et al., 1962). The second-generation fluoroquinolone, CIP, was first used clinically in the mid-1980s, and from 1993, it was used to treat uncomplicated gonorrhea (CDC, 1993; Davis et al., 1996). Fluoroquinolones target DNA gyrase (topoisomerase II) and topoisomerase IV, the enzymes involved in the DNA replication process (Drlica and Zhao, 1997; Larkin et al., 2007). During DNA replication, DNA gyrase catalyzes the unwinding of DNA molecules and topoisomerase IV decatenates daughter chromosomes following DNA replication (Reece and Maxwell, 1991; Kato et al., 1992; Zechiedrich and Cozzarelli, 1995; Zechiedrich et al., 1997). DNA gyrase is comprised of two subunits, GyrA and GyrB (molecular mass ≈96 and 88 kDa, respectively), which are homologous to the C and E subunits of topoisomerase IV, designated as ParC (molecular mass ≈88 kDa), and ParE (molecular mass ≈70 kDa), respectively. Ciprofloxacin resistance in *N. gonorrhoeae* is caused by point mutations in the quinolone resistance-determining region (QRDR) of *gyrA* and *parC*, which results in amino acid substitutions that alter the target protein structure, thereby reducing the fluoroquinolone-target enzyme binding affinity, leading to resistance (Belland et al., 1994; Piddock, 1999; Alcalá et al., 2003; Bodoev and Il'ina, 2015; Yahara et al., 2020). The majority of the resistance mutations in gonococci are located in the QRDR regions, in codons 67–106 in GyrA and 56–108 in ParC (Belland et al., 1994). The known resistant-associated mutations (RAMs) include substitutions at amino acid positions S91 and D95 in GyrA and G85, D86, S87, S88, Q91, and R116 in ParC (Piddock, 1999; Yang et al., 2006; Bodoev and Il'ina, 2015). Studies have shown that GyrA is the primary target and that ParC is the secondary target for fluoroquinolones in *N. gonorrhoeae* (Wittmann et al., 1973; Belland et al., 1994; Trees et al., 1999; Tanaka et al., 2000; Shultz et al., 2001; Lindbäck et al., 2002).

The genus *Neisseria* includes several commensals and two pathogens—*N. gonorrhoeae* and *Neisseria meningitidis*. *Neisseria* species are naturally competent and transfer DNA to one another (Catlin, 1960; Higashi et al., 2011). HGT of *penA*, *mtrCDE*, *rpsE*, and *rplD* from commensal *Neisseria* has been shown to be important in the genesis of resistance to β-lactams and macrolides in the pathogenic *Neisseria* (Bowler et al., 1994; Shafer, 2018; Wadsworth et al., 2018; Yahara et al., 2018; Fiore et al., 2020; Manoharan-Basil et al., 2021). A study from Shanghai has similarly demonstrated that HGT from *gyrA* in commensal *Neisseria* played a critical role in the genesis of fluoroquinolone resistance in *N. meningitidis* (Chen et al., 2020). HGT from commensal streptococci has also been shown to have played a role in the genesis of fluoroquinolone resistance in *Streptococcus pneumoniae* (Irene et al., 1998; Janoir et al., 1999; José et al., 2000). Recently, Yahara et al., 2021 suggested that fluoroquinolone resistance substitutions at GyrA and ParC in *N. gonorrhoeae* isolates have arisen from independent mutations and evidence of HGT in *gyrA/B* and *parC/E* were not available for *N. gonorrhoeae*. It is unknown but essential to establish if HGT has played a role in fluoroquinolone resistance in *N. gonorrhoeae*. If commensal *Neisseria* can serve as a reservoir of resistance for *N. gonorrhoeae*, then addressing the resistance determinants in commensals is important (de Korne-Elenbaas et al., 2021; Kenyon et al., 2021). The above considerations led to the current study, where we used a dataset of 20,047 globally sourced *N. gonorrhoeae* (1928–2020), *N. meningitidis* and *Neisseria* commensals to assess if there was evidence of HGT in the four fluoroquinolone resistance-conferring genes (*gyrA*/*B* and *parC/E*). This led to the identification of HGT in the QRDR regions of *gyrA*, *gyrB*, and *parE* in *N. gonorrhoeae*.

MATERIALS AND METHODS

Dataset for Analysis

The isolates for the analyses were downloaded from the Pathosystems Resource Integration Center V3.6.9 (PATRIC) database (Davis et al., 2020), GenBank,[1] pubMLST (Jolley et al., 2018), and pathogenwatch.[2] Whole-genome sequence (WGS) data from our previous study with an additional 8,388 isolates were used in the current study (Manoharan-Basil et al., 2021). WGS data comprised 20,047 isolates, of which 18,800 were *N. gonorrhoeae*, nine were *N. meningitidis*, and 1,238 were commensal *Neisseria* isolates. **Supplementary Table 1** in **Data Sheet 1** summarizes the organisms used in this study. The European Committee on Antimicrobial Susceptibility Testing breakpoints was used to define ciprofloxacin susceptibility.[3] The minimum inhibitory concentration (MIC) to ciprofloxacin (CIP) ≥16 mg/L, >0.06 mg/L, 0.03 ≤ 0.06, and <0.03 mg/L are classified as being high-level resistant (HLR), resistant (R), intermediate (I), and susceptible (S) to CIP, respectively (Sánchez-Busó et al., 2021). The isolates were categorized into three distinct eras: pre-antibiotic (pre-1950s), golden (1950–1970s), and post-modern (1980-20-first century) following the convention of Golparian et al. (2020).

Gene-By-Gene Analysis, Allelic Profiling, and Recombination Analysis

The analysis was carried out as described in Manoharan-Basil et al. (2021). In brief, WGS (n = 20,047) were analyzed using chewBBACA version 2.8.5 (Silva et al., 2018), followed by recombination analysis using Recombination Detection Program (RDP4) program version 4.100 (Martin et al., 2015). Firstly, a training file was created from the complete genome of *N. gonorrhoeae* FA1090 using Prodigal and was used in subsequent

[1]https://www.ncbi.nlm.nih.gov/genbank/
[2]https://pathogen.watch/ngonorrhoeae
[3]http://www.eucast.org/clinical_breakpoints/

steps (Hyatt et al., 2010). Secondly, a study-specific *Neisseria* scheme was created from two complete *Neisseria gonorrhoeae* genomes (FA1090 and MS11). Thirdly, a FASTA file for each coding sequence (CDS) was generated, followed by the creation of whole-genome (wg) multi-locus sequence typing (MLST) loci. The core-genome (cg) MLST loci were then extracted from the wgMLST loci and visualized using a grape tree (Zhou et al., 2018). GrapeTree facilitates the clustering of the isolates based on their allelic profiles using a minimum spanning algorithm. The SchemaEvaluator option implemented in chewBBACCA uses MAFFT that allows multiple sequence alignments of the alleles of each locus and constructs neighbor-joining (NJ) tree using ClustalW2 (Katoh et al., 2002; Larkin et al., 2007). Finally, UniProtFinder[4] was used to retrieve the functional information of the CDSs. *GyrA*, *gyrB*, *parC*, and *parE* genes were identified based on the UniProt identifier. The multiple sequence alignments and NJ trees of the above genes were extracted from the schema evaluator. The gene-by-gene analysis was carried out on a server with Intel(R) Xeon(R) Silver 4114 CPU @ 2.20 GHz with 125 Gb RAM and HDD distributed RAID-5 using 20 CPU. The analysis took ~15 days to complete.

The multiple sequence alignment files were imported in MEGAX version 10.1.7, and the CDSs were translated (Kumar et al., 2018). The presence or absence of non-synonymous substitutions in the QRDR regions for each amino acid position was denoted as "0" and "1," respectively. The NJ trees and the corresponding metadata were visualized using Interactive tree of life (iTOL) v6 (Letunic and Bork, 2021) and microreact (Argimón et al., 2016).

Whole-genome sequence of commensal *Neisseria* spp., ($n = 1{,}247$) along with *N. meningitidis* ($n = 9$) and *N. gonorrhoeae* ($n = 18{,}800$) were used in the recombination analysis (**Supplementary Table 1** in **Data Sheet 1**). The nucleotide alignments of each gene were screened for recombination events using the Recombination Detection Program (RDP4) program (Martin et al., 2015). Recombinant events supported by at least two of the seven algorithms available in RDP software: RDP, GENECONV, Bootscan, Maxchi, Chimera, SiSscan, and 3Seq were used with default settings, except the window size was increased to 60 nt in RDP, 120 nt in MaxChi and Chimera, and to 500 in BootScan and SiScan (Lemey et al., 2009; Gomez-Valero et al., 2011; Manoharan-Basil et al., 2021). NJ trees were constructed using 1,000 bootstrap replicates (Salminen et al., 1995). According to the definition in RDP4, the minor parent is the parental sequence that contributes the smaller fraction of the recombinant sequence, while the major parent is the parental sequence that contributes the larger fraction (Martin et al., 2015). Additionally, for the sequences which had the same recombination event, the percentage similarities of the nucleotide were determined between donor and recipient sequences. Furthermore, to identify the RAM harboring donors of the *gyrA* and *parC* in *N. gonorrhoeae* and *Neisseria* commensal isolates, RDP4 analysis was carried out as above and was limited to all *Neisseria* commensal and *N. gonorrhoeae* alleles with known RAMs.

[4]https://www.uniprot.org/

Statistical Analysis

Statistical analyses were performed using JMP®, version 14.0.0 (SAS Institute Inc., Cary, NC, 1989–2021). To evaluate the correlation between multiple pairs of variables (SNPs) and to assess the correlations between phenotypic and genotypic patterns of resistance to ciprofloxacin, Spearman's rank correlation was used (Elhadidy et al., 2020). Geometric mean (GM) MICs were calculated for each allele. A value of p of <0.001 was considered statistically significant.

RESULTS

Characteristics of Isolates Used in the Study

A total of 20,047 isolates from 1928 to 2020 were used in the study and included 18,800 (93.7%) *N. gonorrhoeae*, nine (0.47%) *N. meningitidis*, and 1,247 (6.2%) *Neisseria* commensal isolates. *Neisseria* commensal included the human nasopharyngeal commensals *N. polysaccharea* ($n = 74$), *N. bergeri* ($n = 66$), *N. lactamica* ($n = 699$), *N. cinerea* ($n = 52$), *N. subflava* ($n = 98$), *N. oralis* ($n = 8$), *N. mucosa* ($n = 38$), *N. elongata* ($n = 32$), and *N. bacilliformis* ($n = 8$), along with several species isolated from animals and animal-related source ($n = 163$; **Supplementary Table 1** in **Data Sheet 1**; **Figure 1**).

Out of the 20,047 isolates, 9,796 (49.8%) samples were from 31 European countries, 3,399 (16.9%) from North America, 3,485 (17.3%) from Oceania, 1,750 from Asia (8.7%), 477 (2.37%) from Africa, and 162 (0.8%) from South America (**Figure 2**; **Supplementary Table 2** in **Data Sheet 1**). The country of isolation was not available for 978 (4.8%) isolates.

Antimicrobial Susceptibility of *Neisseria* Isolates

Among the 18,800 *N. gonorrhoeae* isolates, the CIP MIC was available for 12,942 (68.8%) isolates. Out of these isolates, 6,192 (47.8%) of the isolates were CIP resistant (R), 45 (0.34%) isolates were CIP intermediate (I), and 6,705 (51.8%) isolates were susceptible to CIP (**Supplementary Table 2** in **Table 2**). Out of the 1,247 *Neisseria* commensal species, CIP MIC was available for only 14 isolates. Seven of these isolates had MICs >0.06 mg/L.

Characterization of Resistance-Associated Mutations in QRDR Regions of *gyrA*, *gyrB*, *parC*, and *parE* Genes in *Neisseria gonorrhoeae*

Strong positive correlations between CIP MICs and substitutions in GyrA at amino acid positions 91 (S91F, 45.4%) and 95 (D95A, 13.82%; D95G, 28.2%; D95N, 3.04%) and ParC at 85 (G85C, 0.5%; G85D, 0.2%), 86 (D86N, 6.8%), 87 (S87I, 1%; S87N, 3.4%; S87R, 25.7%), 88 (S88P, 2.1%), and 91 (E91G, 6.3%; E91K, 0.7%; E91Q, 0.9%) were observed (**Supplementary Table 3** in **Data Sheet 1**).

In vitro generated zoliflodacin-resistant mutants containing substitutions in GyrB (D429, K450, and S467N) have been found to be associated with increased MICs of zoliflodacin

TABLE 1 | Distribution of 5 predominant ST-types of N. gonorrhoeae isolates, country of isolation and their known resistance-associated mutations.

0	CIP resistance	Africa	Asia	Europe	North America	Oceania	South America	NA	Total no. of isolates
1579			56	251	197	219	1	19	743
	HLR		51	4	4			2	61
	I			1	2				3
	R		4	6	21	5		1	37
	S			91	78	214	1	16	400
	NA		1	149	92				242
1901			152	1073	1007	126		284	2677
	HLR		75	322	567	31	24	182	1201
	R		41	102	235	86	11	76	551
	S			11	1			1	13
	NA		36	638	204	9		25	912
7359			20	150		483			653
	R			1					1
	S		20	126		296			442
	NA			23		187			210
7363		1	237	446	85	252	1	28	1050
	HLR		82	164	15	2	1	10	274
	R		47	88	12	230		9	386
	S	1	2	12		13			28
	NA		106	182	58	7		9	362
9363				597	548	418	9	55	1627
	HLR			16	11			1	28
	I			7	4				11
	R			6	29	4		1	40
	S			287	335	412	9	52	1095
	NA			281	169	2		1	453

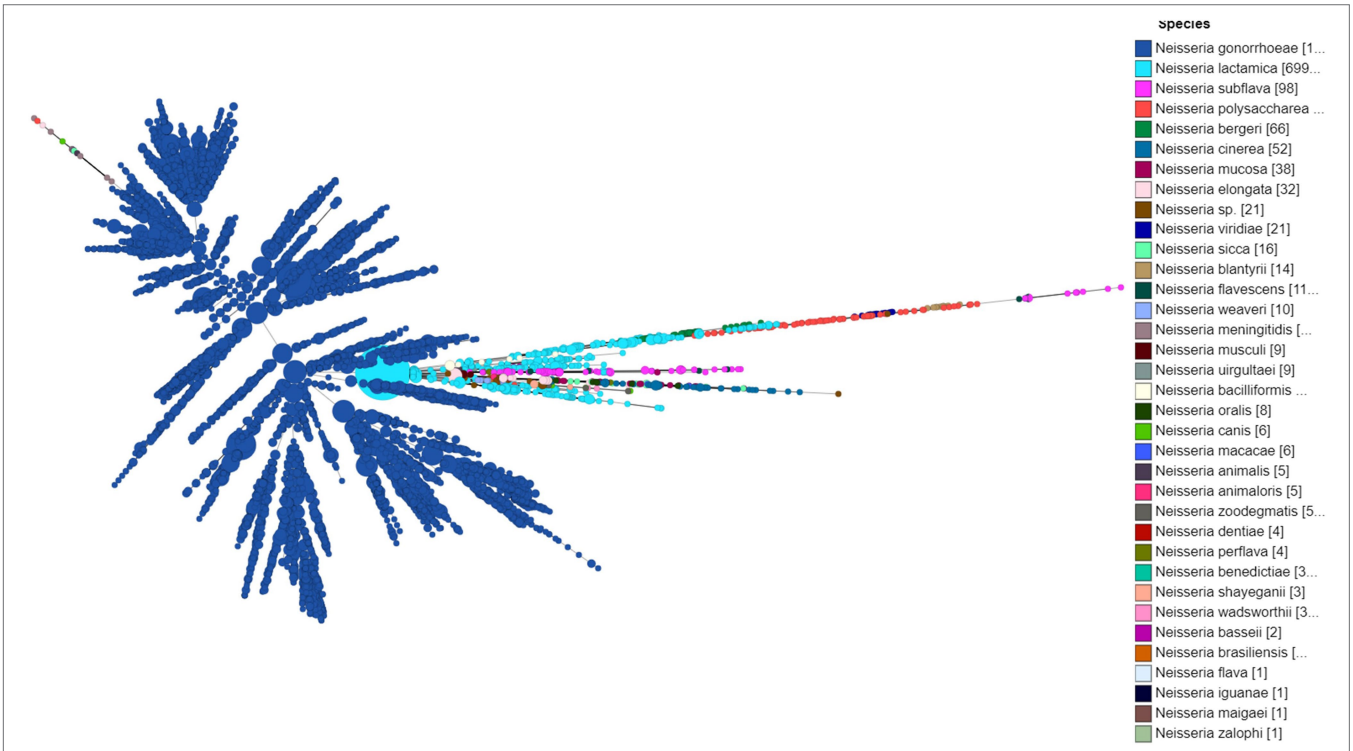


FIGURE 1 | Minimum spanning tree comparing core-genome allelic profiles in association with the *Neisseria* species used in the study. Numbers in brackets refer to the number of isolates. Isolates are displayed as circles. The size of each circle indicates the number of isolates of this particular type.

(Alm et al., 2015; Foerster et al., 2015, 2019). In the current study, substitutions in GyrB-D429V were present in a single Japanese isolate (PubMLST ID-31741; Pathogenwatch ID-ERR363582; Adamson et al., 2021) along with GyrA-S91F

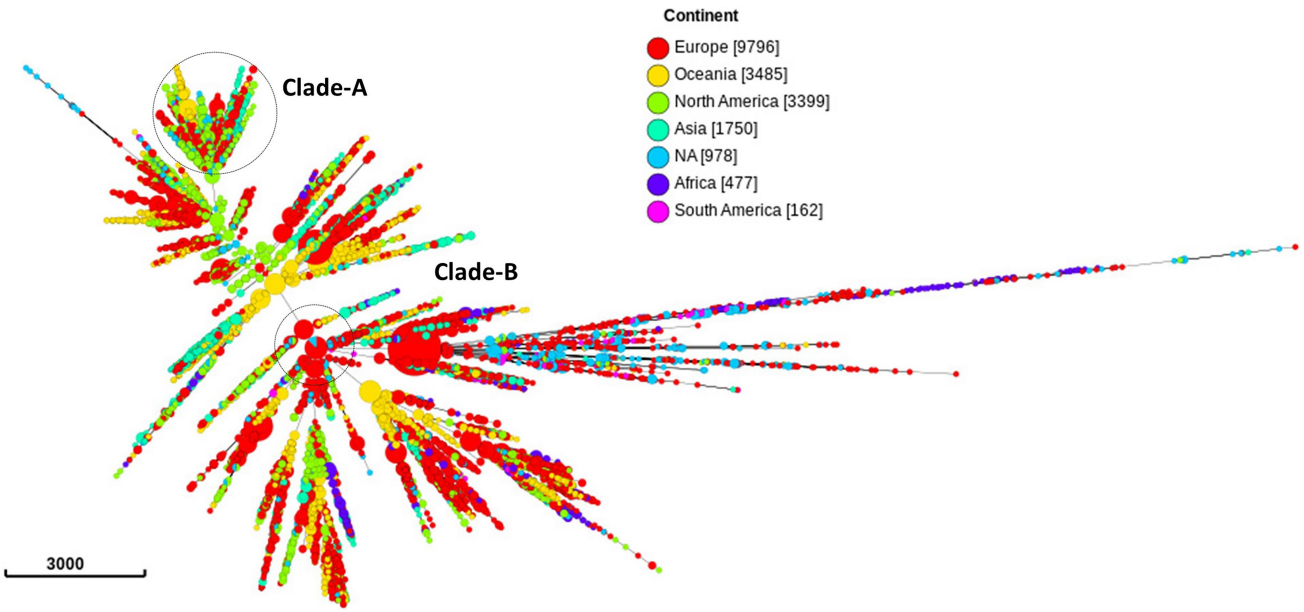


FIGURE 2 | Minimum spanning tree comparing core-genome allelic profiles in association with continents. Numbers in brackets refer to the number of isolates. Isolates are displayed as circles. The size of each circle indicates the number of isolates of this particular type. Clade A, B of ST1901 are represented as dotted circles.

and ParC-S87R substitutions. Fifty isolates had the GyrB-S467N substitution [allele 293 ($n = 12$) and allele-446 ($n = 38$); Pathogenwatch ($n = 24$) and PubMLST ($n = 26$)] along with mutations in *gyrA* and or *parC*. These isolates were from China ($n = 1$, ST7367), Japan ($n = 2$, ST7363), Norway [$n = 11$; ST1925 ($n = 10$), ST7363 ($n = 1$)], and Vietnam [$n = 36$; ST7373 ($n = 34$), ST1925 ($n = 2$)]. None of the isolates had the GyrB-K450 substitution.

Except 27 isolates, all high-level fluoroquinolone (CIP MICs ≥16 mg/L) resistant gonococcal isolates, $n = 2,384$ (12.7%) had known substitutions at GyrA-91 (S91F), GyrA-92 (A92P), GyrA-95 (D95A/G/N), and ParC-85 (G85C), 86 (D86N), 87 (S87I/N/R/Y), 88 (S88P), and 91 (E91G/K) positions (**Figures 3–5**; **Supplementary Table 2** in **Table 2**). Twenty-seven isolates (1.1%; GM MIC = 22.3 mg/L), 14 (0.5%; GM MIC = 27.5 mg/L), 2,040 (85%; GM MIC = 21.1 mg/L), 273 (11.4%), and 3 (0.1%) had one, two, three, four, and five RAMs, respectively (**Supplementary Table 1** in **Table 1**). A total of 3,808 (20.2%) *N. gonorrhoeae* isolates had low level fluoroquinolone resistance (CIP MICs >0.06 and <16 mg/L). Out of these isolates, 79 isolates (2%, GM MIC = 1.53 mg/L) had no known RAMs (**Figures 3–5**; **Supplementary Table 2** in **Table 2**). Eighty-nine (2.3%; GM MIC = 0.64 mg/L), 302 (7.9%; GM MIC = 27.5 mg/L), 3,017 (79.2%; GM MIC = 21.18 mg/L), 326 (8.5%), and 5 (0.1%) isolates had one, two, three, four, and five RAMs, respectively (**Supplementary Table 1** in **Table 1**). Out of 6,705 (35.7%) susceptible gonococcal isolates, 6,620 (98.7%) had no GyrA or ParC substitutions, whereas 46, 9, 33, and 2 isolates had one, two, three, and four RAMs, respectively (**Supplementary Table 1** in **Table 1**). Seven *Neisseria* commensals (0.08%) with ciprofloxacin MICs >0.06 mg/L, had GyrA-S91I/T/V or GyrA-D95N substitutions.

The different combinations of RAMs and the number of isolates are presented in **Figures 3–5** and **Supplementary Figure 1**.

Genetic Structure and Geographical Dispersal of Fluoroquinolone Resistance in *Neisseria gonorrhoeae*

A total of 1,323 loci were defined as core to the *Neisseria* genome in this dataset. Allelic designations were defined for approximately 95% of the core-genome. A total of 583 different sequence types (STs) characterized the available population. The most frequent ST was ST1901 ($n = 2,677$) followed by ST9363 ($n = 1,627$), ST7363 ($n = 1,054$), ST1579 ($n = 743$), and ST7359 ($n = 653$; **Figure 6**; **Supplementary Table 2** in **Table 2**). MLST types were not available for 2,424 isolates, and 93 isolates were classified as new ST types. Of the 2,384 CIP HLR isolates detected, half of the isolates [1,201 (50.4%)] belonged to ST1901, followed by ST7363 (11.49%) and ST1579 (2.5%). ST types were not available for 231 (9.6%) isolates. The total number of HLR, I, R, and S isolates and their prevalence in each continent is listed in **Supplementary Table 4** in **Data Sheet 1**.

The three most prevalent ST types (>1,000 isolates) were further characterized. ST1901 was split into two clades, referred to as Clade A and Clade B (**Figure 6**). Isolates of clade A were predominately from North America, while clade B isolates were predominately from Europe (**Figure 2**). Both clades had substitutions at GyrA-S91F ($n = 2,616$, **Figure 4A**;

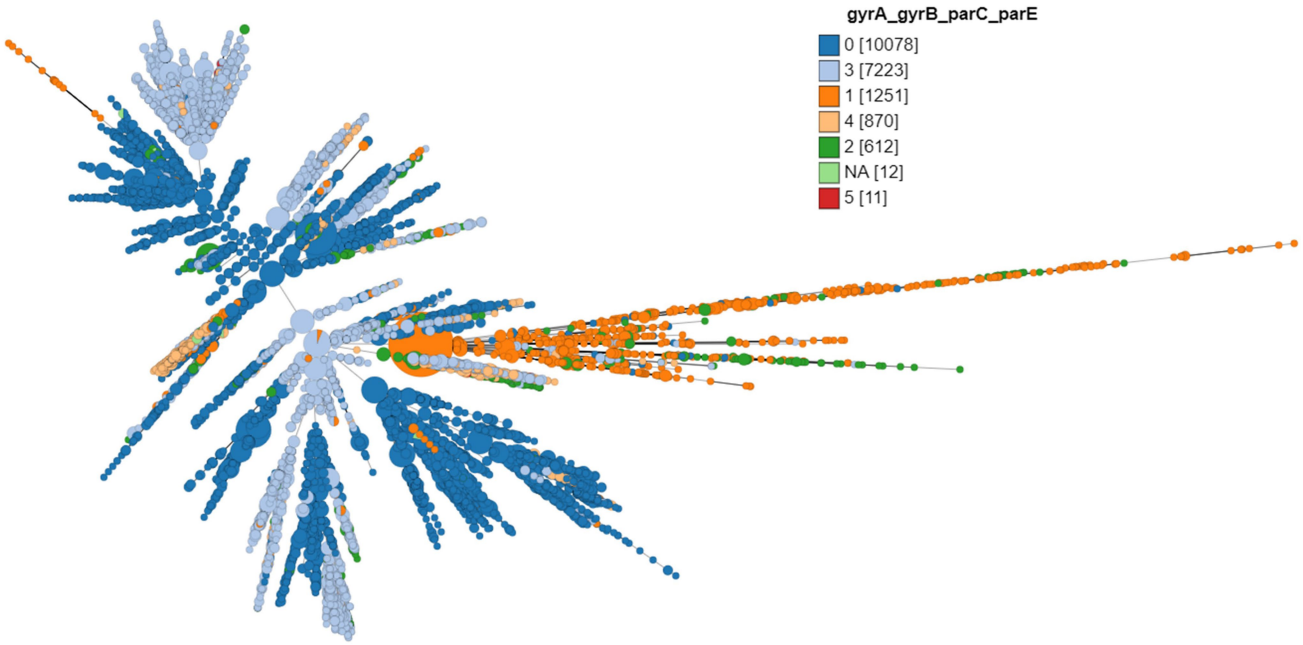


FIGURE 3 | Minimum spanning tree comparing core-genome allelic profiles with the number of substitutions in GyrA, GyrB, ParC, and ParE. Numbers in brackets refer to the number of isolates. Isolates are displayed as circles. The size of each circle indicates the number of isolates of this particular type.

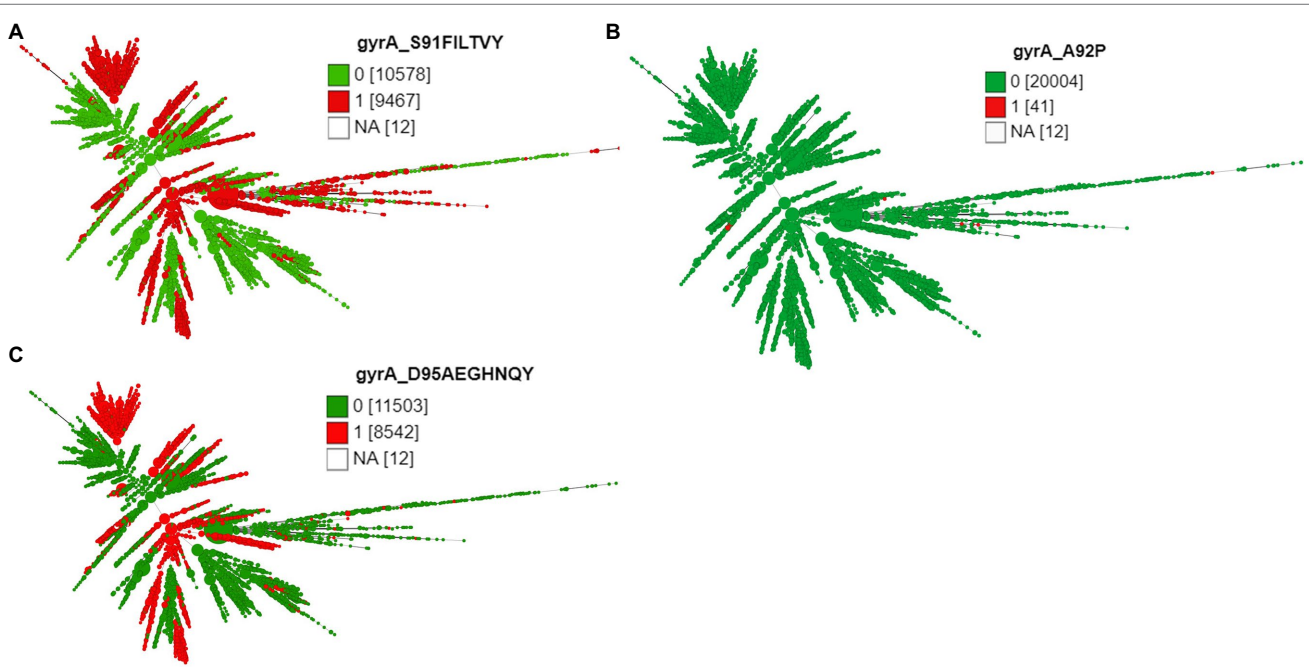


FIGURE 4 | Minimum spanning tree comparing core-genome allelic profiles with GyrA substitutions at amino acid **(A)** position 91, **(B)** position 92, and **(C)** position 95. Numbers in brackets refer to the number of isolates. 1 and 0 denotes the presence and absence of RAMs, respectively. Isolates are displayed as circles. The size of each circle indicates the number of isolates of this particular type.

Supplementary Table 2 in **Table 2**), GyrA-D95G ($n = 2,585$, **Figure 4C**), and ParC-S87R ($n = 2,663$, **Figure 5C**). Around 0.4, 20.5, and 44% of isolates belonging to ST1901 were S, R, and HLR, respectively. About 34% of isolates had no MIC information (**Table 1**).

ST9363 isolates were distributed across Europe (36.69%), North America (33.68%), and Oceania (25.69%; **Table 1**; **Figure 2**). The first HLR isolate was found in North America in 2011. About 67.3, 0.6, 2.4, and 1.7% of isolates were S, I, R, and HLR, respectively. About 27.8% of isolates had no

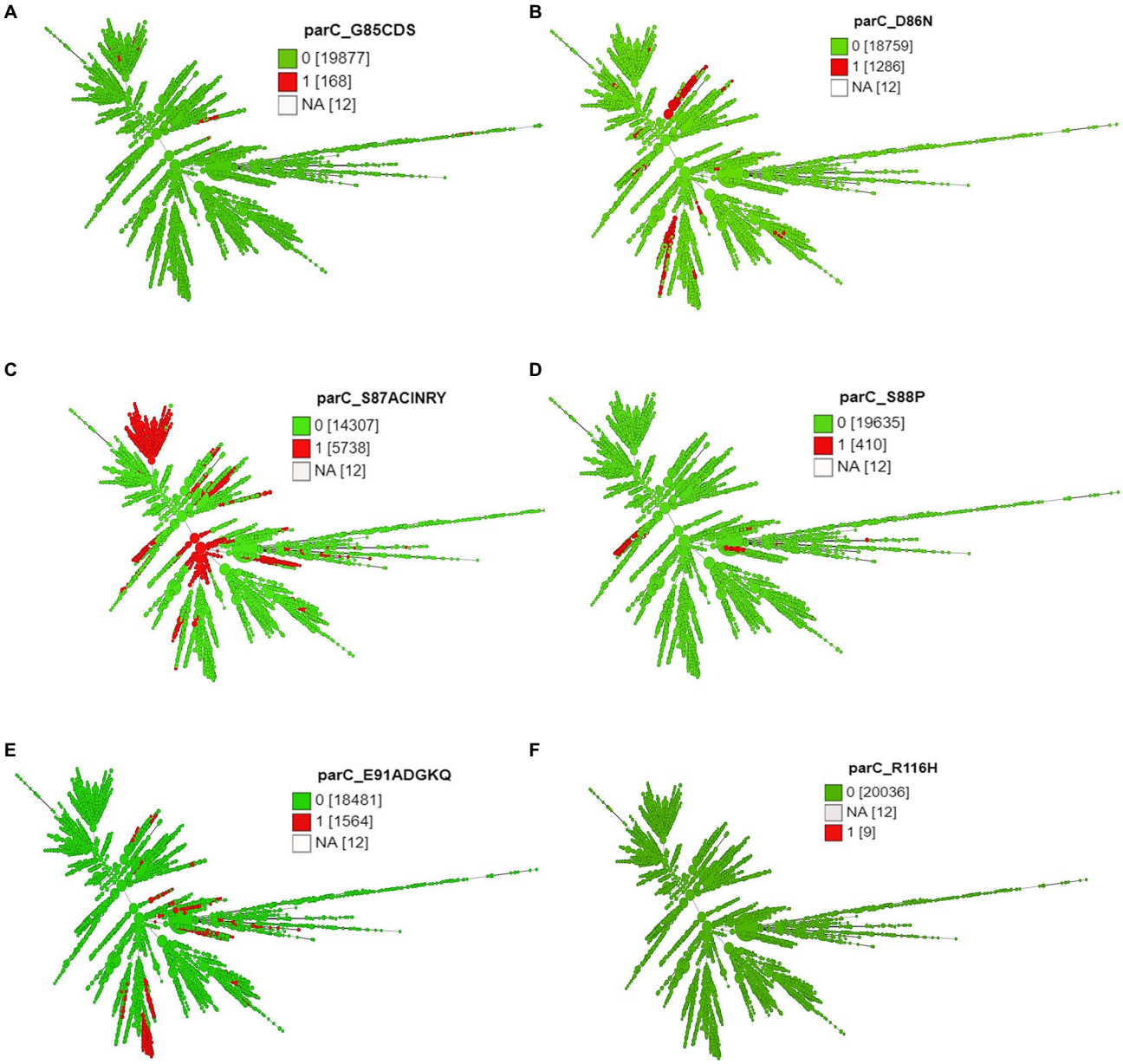


FIGURE 5 | Minimum spanning tree comparing core-genome allelic profiles with ParC substitutions at amino acid **(A)** position 85, **(B)** position 86, **(C)** position 87, **(D)** position 88, **(E)** position 91, and **(F)** position 116. Numbers in brackets refer to the number of isolates. 1 and 0 denotes the presence and absence of RAMs, respectively. Isolates are displayed as circles. The size of each circle indicates the number of isolates of this particular type.

MIC information. The resistant isolates had one or more of the following substitutions: GyrA-S91F ($n = 78$, **Figure 4A**), GyrA-D95A/G ($n = 76/2$, **Figure 4C**), ParC-D86N ($n = 71$, **Figure 5B**), and the ParC-S87R ($n = 10$, **Figure 5C**).

At least three clades for ST7363 were observed (**Figure 6**). Isolates that belonged to ST7359 were mostly from Oceania (**Figure 2**). Out of the 653 isolates, 442 were susceptible, and MIC was not available for 210 isolates. Only one isolate from Europe (Ireland) was CIP resistant (MIC-2 mg/L) with no known RAMs. Further details pertaining to the distribution of MLST, CIP MICs, and the prevalence in each continent are provided in **Table 1** and **Supplementary Table 4** in **Data Sheet 1**.

The Time of Emergence of Fluoroquinolone Resistance-Associated Mutations in *Neisseria gonorrhoeae*

Substitution at position 91 (S91T) of GyrA was first observed in one of two of the oldest isolates in the collection obtained in 1928. The MICs were not available for the pre-antibiotic and the golden era isolates. GyrA S91F, which leads to high-level resistance to CIP, was first observed in 1992 (post-modern

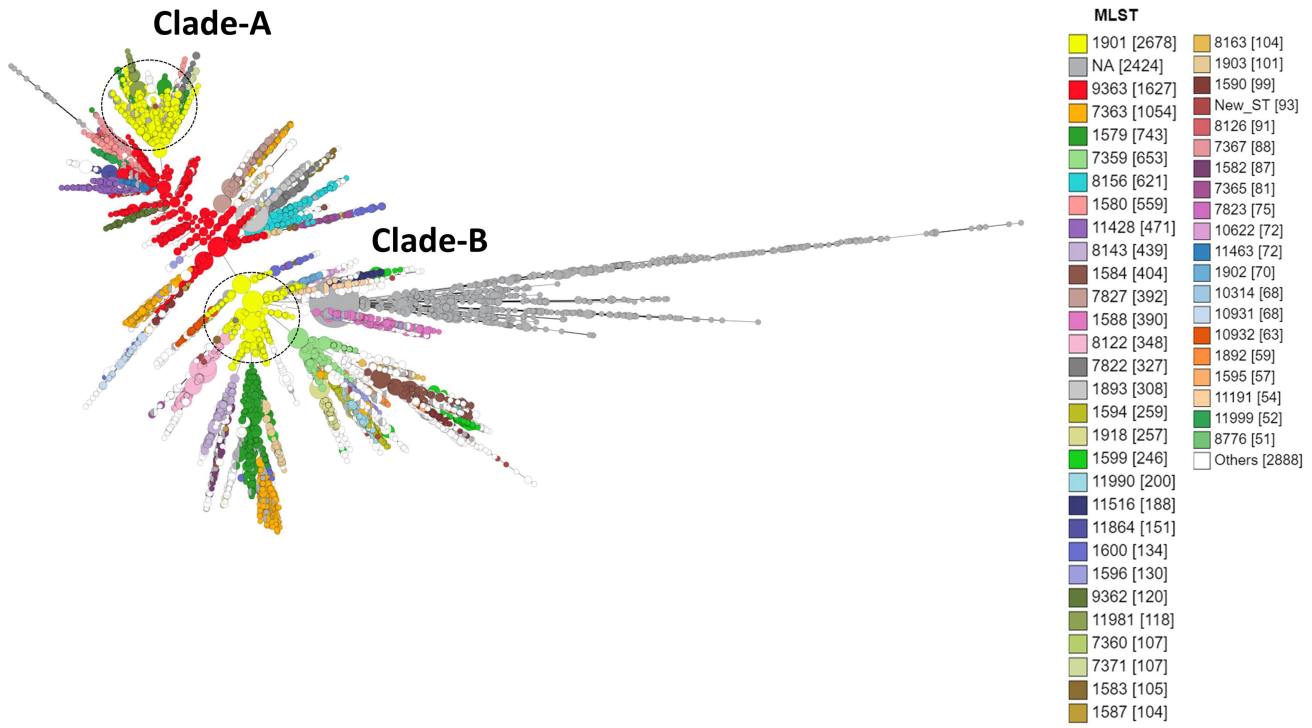


FIGURE 6 | Minimum spanning tree comparing core-genome allelic profiles with MLST resulting in isolates with similar allelic profiles forming clusters. Isolates are displayed as circles. The size of each circle indicates the number of isolates of this particular type. Numbers in brackets refer to the number of isolates. Clade A, B of ST1901 are represented as dotted circles.

era). Four isolates were available from the same year. Two isolates were from the Philippines and had GyrA-S91F along with additional substitutions, and two isolates had no RAMs. Out of the two isolates with RAMs, one isolate belonged to ST1901 (D95G; ParC-S87R), and the other to ST7367 (D95G). Substitutions at position 95 of GyrA (D95G/N) and positions 86, 87, 88, and 91 of ParC were first observed in 1996. A substitution in position 85 of ParC was first observed in 1998 (**Figure 7**; **Supplementary Table 6** in **Data Sheet 1**).

Prevalence of HGT in QRDR Region of *gyrA*, *gyrB*, *parC*, and *parE*

The lengths of *gyrA*, *gyrB*, *parC*, and *parE* of the reference isolate FA1090 were 2,751, 2,391, 2,307, and 1,986 bp, respectively. The length of the alleles ranged between 2,454–3,054 bp for *gyrA*, 2,388–2,644 bp for *gyrB*, 2088–2,538 bp for *parC*, and 1,629–2,151 bp for *parC*. Multiple alignment files of each gene were trimmed to the gene length of the FA1090 isolate and were used in the recombination analysis.

About 688, 631, 898, and 595 alleles were found for *gyrA*, *gyrB*, *parC*, and *parE* genes, respectively (**Figures 8A–D**; **Supplementary Tables 6A–D** in **Data Sheet 1**). Putative HGT events were predicted in all the genes (**Figures 9A–D**). Unique recombination events were supported by at least two out of seven detection methods. A total of 63, 8, 80, and 67 recombination events were identified for *gyrA*, *gyrB*, *parC*, and *parE*, respectively. For the sequences that included only the RAMs in *gyrA* and *parC*, a total of 58 recombination events were identified for *gyrA* (**Supplementary Table 7E** in **Data Sheet 1**) and 16 for *parC* gene (**Supplementary Table 7F** in **Data Sheet 1**). Each recombination event was checked manually. Examples of recombinant events with known RAMs in QRDR regions of *gyrA* and *parC* and in QRDR regions of *gyrB* and *parE* are presented below.

gyrA

For *gyrA*, the QRDR region is present between 169 and 432 bp. Putative recombination event 32 in the QRDR region with *N. gonorrhoeae* as a recombinant was identified. The event was supported by five methods (RDP, GENECONV, Bootscan, Maxchi, and Chimaera). *Neisseria gonorrhoeae* (allele—488; $n = 2$, GM MIC-0.02 mg/L, ST11702, Year-1940) was the recombinant, and evidence of the same recombinant event was present in four other alleles (allele-93, $n = 3$, ST1175, Year-2010–2012; allele-491; $n = 2$, ST1688, Year-1928; allele-496, $n = 2$, ST11703, Year-1951; and allele-508, $n = 2$, ST11708, Year-1951). The corresponding minor parent was *N. lactamica* (allele-270; $n = 1$; recombinant region—197–396 nt), and major parent was *N. cinerea* (allele-12; $n = 1$; recombinant region—1–199 and 397–3,168 nt; **Figure 9A**; **Supplementary Table 7A** in **Data Sheet 1**; **Supplementary Figure 2**). The minor parent *N. lactamica* and the recombinants had the S91T substitution (**Supplementary Table 6A** in **Data Sheet 1**).

Era	Year	Total No. of Isolates	No.of isolates with RAMs		Genotype	
Pre-antibiotic	1928	2	1	--->	gyrA (S91T) *	
	1940	10	1	--->	gyrA (S91T)	
Golden	1951	6	2	--->	gyrA (S91T)	
	1952	1	1	--->	gyrB (M436K) *	
	1960	2	1	--->	gyrB (M436K)	
	1962	4	2	--->	gyrB (M436K), parC (E91D) *	Quinolone - Nalidixic acid (UTI)
	1973	2	1	--->	gyrB (M436K)	
	1974	1	1	--->	gyrB (M436K)	
	1977	1	2	--->	gyrA (S91T)	
	1979	12	3	--->	gyrA (S91T)	
Post-modern	1980	3	1	--->	gyrB (M436K)	Quinolone – Ciprofloxacin (UTI)
	1982	6	2	--->	gyrB (M436K)	
	1983	3	2	--->	gyrA (S91T)	
	1984	4	2	--->	gyrB (M436K)	
	1985	3	1	--->	gyrB (M436K)	
	1987	2	1	--->	gyrB (M436K)	
	1989	8	4	--->	gyrA (S91T)	
	1991	6	2	--->	gyrA (S91T)	
	1992	11	9	--->	gyrA (S91F/T/Y) (D95G), gyrB (M436K), parC (S87R) *	
	1993	7	3	--->	gyrA (S91F), gyrB (M436K)	Fluoroquinolone – Ciprofloxacin Uncomplicated gonorrhea
	1996	44	17	--->	gyrA (S91F/T/Y) (D95G/N), parC (S87R), parC (D86N) (S88P) (R116H) *	
	1997	114	46	--->	above mutations & parC E91G, parC E91G *	
	1998	110	74	--->	above mutations & parC (G85C) (S87N/R), parC (E91K) *	
	1999	44	26		"	
	2000	27	42		"	
	----- 2020	5	7			

FIGURE 7 | Time of emergence of ciprofloxacin resistance-associated mutations. Asterisks denote the first occurrence of the mutation.

Recombination events were also found in the commensal *Neisseria*. In event 5, for example, *N. polysaccharea* (allele—361; $n = 1$) was the recombinant, and the sequences with the same recombinant event were present in 10 isolates belonging to five alleles; *N. bergeri* (alleles- 28, 214, 215, and 223; $n = 9$) and *N. oralis* (allele-29, $n = 1$). The corresponding minor parent was *N. cinerea* (allele-61; $n = 1$; recombinant region—1–172 and 2,946–3,168 nt) and major parent was *N. elongata* (allele-637;

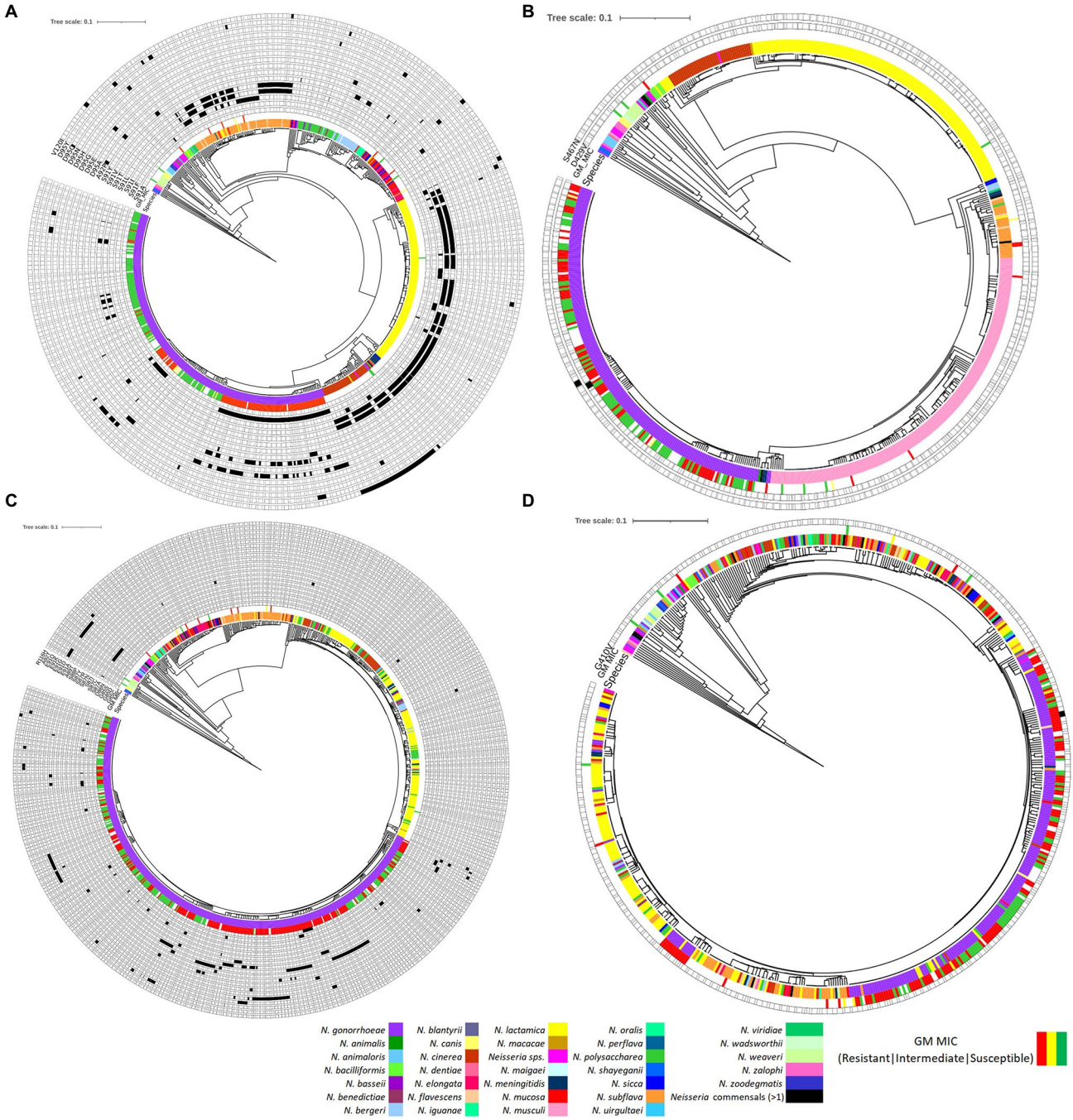


FIGURE 8 | Phylogenetic trees of **(A)** *gyrA*, **(B)** *gyrB*, **(C)** *parC*, and **(D)** *parE* based on *Neisseria* spp. Figure was generated by iTOL using cgMLST profile data determined by the chewBACCA software. Purple node denotes *N. gonorrhoeae*, and the rest denotes *Neisseria* commensals. Filled and outlined black rectangle denotes the presence and absence of non-synonymous SNPs, respectively. The geometric mean (GM) ciprofloxacin MIC SIR profiles are denoted as green, yellow, and red, respectively.

$n = 2$; GM MIC-0.25 mg/L) recombinant region—173–2,945 nt (**Supplementary Table 7A** in **Data Sheet 1**). The major parent *N. elongata* (allele-637), the recombinant *N. polysaccharea* (allele—361), and isolates with the same recombinant ($n = 10$) had the S91V substitution (**Supplementary Table 6A** in **Data Sheet 1**). The event was supported by all seven methods.

gyrB

The QRDR region of *gyrB* is between 1,255 and 1,488 bp, which is also the zoliflodacin resistance-conferring region (**Figure 9B**). Recombination event 3 was supported by all seven methods (RDP, GENECONV, Bootscan, Maxchi, Chimera, SiSscan, and 3Seq; **Figure 9B**) and *N. meningitidis* ($n = 1$; allele-577) was the recombinant. The same recombination event was present in five

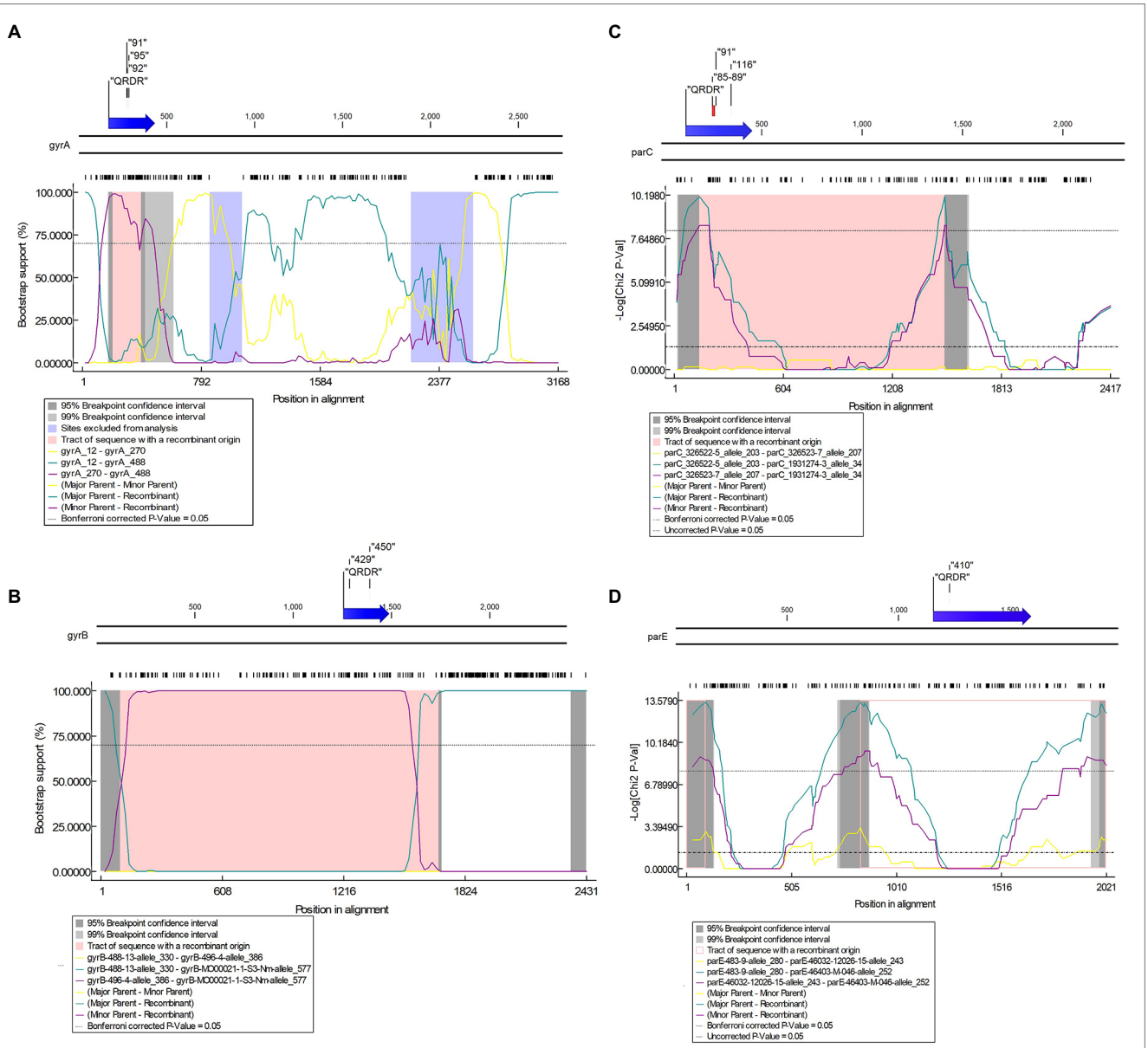


FIGURE 9 | Potential putative recombination events between *Neisseria* strains. **(A)** Recombination event 3 in *gyrA* between *N. gonorrhoeae* (recipient) and *N. lactamica* (donor); **(B)** Recombination event 3 in *gyrB* between *N. gonorrhoeae*/*N. meningitidis* (recipient) and *N. macacae* (donor); **(C)** Recombination event 21 in *parC* between *Neisseria* sp. (recipient) and *N. zoodegmatis* (donor); **(D)** Recombination event 19 in *parE* between *N. gonorrhoeae* (recipient) and *N. mucosa* (donor). Top—The functional map of *N. gonorrhoeae* reference strain FA1090 with QRDR and known resistance-associated mutations (RAMs) are shown within inverted commas. Bottom—Bootscan **(A,B)** and maxchi **(C,D)** plots generated using RDP4 software are depicted.

alleles, including *N. gonorrhoeae* ($n = 13$, allele-472, 518, and 562) and *N. meningitidis* ($n = 2$; allele-544 and 567). The minor and major parents were *N. macacae* (allele-386; $n = 1$; region—96–1,690 nt) and *N. mucosa* (allele-330; $n = 1$; region—1–95 and 1,691–2,431 nt), respectively (**Figure 9B**; **Supplementary Tables 5**, **7B** in **Data Sheet 1**). The recipient and the donor had no known RAMs (**Supplementary Table 6B** in **Data Sheet 1**).

parC

The QRDR region lies between 121 and 452 bp. Recombination event 21 took place in the QRDR region and involved the transmission of known RAMs between commensal *Neisseria*. The recombinant was *Neisseria* sp. (allele—34; $n = 2$). *Neisseria animaloris* (allele—203; $n = 1$; recombinant region—1–141 and 1,499–2,417 nt) and *N. zoodegmatis*. (allele—207; $n = 1$; recombinant region—142–1,498 nt) were the major and minor parents, respectively (**Figure 9C**). The same event was seen in another allele (allele-35, $n = 2$). The minor parent and the recombinant had the S87A and E91A RAMs (**Supplementary Table 6C** in **Data Sheet 1**). The event was supported by three methods (Maxchi, Chimera, and 3Seq).

parE

The QRDR region lies between ~1,140 and 1,600 bp. Recombinant event 19 for *parE* was supported by three methods (Maxchi, Chimera, and 3Seq). The recombinant was *N. gonorrhoeae* (allele-252; $n = 1$). The corresponding major parent was *N. cinerea* (allele 280, $n = 1$; recombinant region—92–837 nt), and the unknown parent was inferred from *N. mucosa* (allele-243; $n = 3$; recombinant region—1–91 and 838–2,021 nt; **Figure 9D**). The recipient and the donor had no known RAMs (**Supplementary Table 6D** in **Data Sheet 1**).

DISCUSSION

The large and diverse collection of *Neisseria* isolates used in this study enabled us to investigate the temporal and spatial patterning of fluoroquinolone RAMs as well as the evidence of HGT in *gyrA*, *gyrB*, *parC*, and *parE* genes in *N. gonorrhoeae* and related *Neisseria* spp.

In accordance with previous studies, we found that the global spread of ciprofloxacin-resistant *N. gonorrhoeae* is associated with three STs 1,901, 9,363, and 7,363 being responsible for 65% of all high-level resistance to ciprofloxacin (Osnes et al., 2020; Reimche et al., 2021). ST1901, with substitutions in GyrA-S91F, D95G, and ParC-S87R, was by itself responsible for over half of all high-level resistance. Osnes et al. have recently demonstrated that this ST emerged in East Asia before spreading globally between 1987 and 1996 (Osnes et al., 2021). Its emergence and spread were facilitated by the increasing use of fluoroquinolones as a treatment for gonorrhea at this time (Osnes et al., 2020). The widespread use of fluoroquinolones in the general population may also have played a role (Kenyon et al., 2018a). Notably, no global distribution of ST1901 clones with substitutions in GyrA-S91Y and ParC-R116H during the same period from East Asia was observed.

The GyrA-S91T substitution was first reported in 1928 in *N. gonorrhoeae* and cannot be plausibly linked with antimicrobial pressure. GyrA-S91T is not associated with reduced susceptibility to ciprofloxacin (Golparian et al., 2020). Notably, two isolates from 1928 had different amino acid substitutions at position 91. One isolate belonging to ST11688 had the threonine (T) amino acid at position 91, whereas the other isolate from ST11680 had the serine (S) at position 91. Interestingly, the meningococci have the threonine (T) substitution at position 91 (Hong et al., 2013). Resolving the parental isolate of gonococci could resolve the ST types of gonococci, especially the two clades of ST1901. Further analysis pertaining to the above was not carried out as this was beyond the scope of the present study. The next mutation to occur in the four genes under consideration was a substitution in ParC-E91D in 1962 in *N. canis*. This occurred at the time when the first-generation quinolone, nalidixic acid, was introduced (Lesher et al., 1962).

Our analyses suggest that 199 bp of the QRDR region in *gyrA* with GyrA-S91T substitution was acquired from *N. lactamica*. Fourteen isolates of *N. gonorrhoeae* belonging to five alleles had obtained this section of the QRDR region, including the GyrA-S91T. All the isolates belonged to the pre-antibiotic/golden era except three isolates from allele-93 that belonged to the post-modern era with a GM MIC of 0.006 mg/L. Evidence of HGT in QRDR regions of *Neisseria* spp. and mutation harboring commensals with substitutions in known amino acid positions, GyrA-S91, ParC-S87, and ParC-E91 were identified. The recombinants with GyrA-S91V and ParC-S87A, ParC-E91A were *N. polysaccharea* and *Neisseria* sp., respectively. The mutation harboring commensal of GyrA-S91V in the recombinant *N. polysaccharea* ($n = 1$, MIC-NA) was *N. elongata* ($n = 2$, GM MIC—0.25 mg/L) and ParC-S87A, E91A in the recombinant *Neisseria* sp. ($n = 2$, MIC-NA) was *N. zoodegmatis* ($n = 1$, MIC-NA). Twenty-one *Neisseria* commensal isolates had the GyrA-S91V substitution, but only one isolate, *N. elongata* had the data for ciprofloxacin susceptibility (MIC 0.25 mg/L). It was therefore not possible to determine if the mutation is associated with reduced susceptibility to ciprofloxacin. The same implies to the substitutions in ParC.

The analysis of whole-genome sequences identified non-synonymous mutations in the GyrB-D429V and S467N, that is, where zoliflodacin first- or second-step resistance mutations have been previously selected *in vitro* (Alm et al., 2015; Foerster et al., 2015, 2019). Substitutions in GyrB-D429V was present in one isolate belonging to ST7363 (Adamson et al., 2021; Unemo et al., 2021). GyrB-S467N substitution was present in 50 isolates and belonged to ST7367 ($n = 1$), ST7363 ($n = 37$), and ST1925 ($n = 12$). Evidence of HGT in the QRDR region of *gyrB* was identified, and the commensal donor was *N. macacae*.

Caveats of our study include that only a few early isolates were available, and very few isolates from some continents, such as Africa and South America, were available. The study included only nine *N. meningitidis* isolates. The probability of detecting HGT events could have been maximized using a greater number of *N. meningitidis*. Moreover, MIC values were not available for many commensal isolates. Some of the Pathogenwatch MICs are estimated based on genotype using 485.toml.[5] These estimated MICs may differ slightly from MICs obtained *via* phenotypic testing. We were also not able to identify the commensal donors for *parC* RAMs. This could be due to the paucity of commensal isolates with known RAMs. We did not include other genetic mutations, such as the *norM* promoter mutation, which can lead to increased fluoroquinolone resistance (Rouquette-Loughlin et al., 2003). Finally, the HGT events were not confirmed experimentally.

Studies have shown that chromosomal *gyrA* and *parC* mutations from commensal *Neisseria* have been responsible for most fluoroquinolone resistance in meningococci (Chen et al., 2020; Shen and Chen, 2020). With the current dataset, we showed evidence of HGT in genes *gyrA*/*B* and *parC*/*E* and found that HGT played a lesser role in acquiring fluoroquinolone resistance mutations in gonococci. Nevertheless, our findings add to the growing evidence base

[5]https://gitlab.com/cgps/pathogenwatch/amr-libraries/blob/0.0.17/485.toml

that commensal *Neisseria* species act as a reservoir of antimicrobial resistance that can be taken up by the pathogenic *Neisseria* (Bowler et al., 1994; Shafer, 2018; Yahara et al., 2018; Fiore et al., 2020; Goytia et al., 2021; Manoharan-Basil et al., 2021). The study highlights the significant role that HGT can play in transferring QRDR regions to *N. gonorrhoeae*. Zoliflodacin is a promising new drug to treat *N. gonorrhoeae* (Foerster et al., 2019; Bradford et al., 2020; Adamson et al., 2021; Kenyon et al., 2021). Our finding that zoliflodacin resistance-conferring region of *gyrB* in *N. gonorrhoeae* can be taken up from *N. macacae* has important consequences. Extensive use of zoliflodacin could lead to resistance in commensal *Neisseria*, which could then be transferred to *N. gonorrhoeae via* transformation. This provides an additional rationale to conduct surveillance of antimicrobial susceptibility in *Neisseria* commensals to prevent the emergence of AMR in the pathogenic Neisseria (Kenyon et al., 2018b; Dong et al., 2020; Fiore et al., 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SM-B and CK conceptualized the study, interpreted the data, performed the statistical analysis, and wrote the first draft. SM-B was responsible for data collection and bioinformatic analysis. All authors read the final draft and approved the submitted version.

FUNDING

The study was funded by SOFI 2021 grant—"PReventing the Emergence of untreatable STIs *via* radical Prevention" (PRESTIP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.793612/full#supplementary-material

Supplementary Presentation 2 | Alignment of *GyrA* alleles-1 (reference, FA1090), 3, 496, 488, 491, 508 from *N. gonorrhoeae* and allele-270 from *N. lactamica* (donor). The quinolone resistance determining region (QRDR), the 199 bp of the recombinant region and the resistance associated mutations (RAM) are depicted in red, green and blue arrows, respectively. Amino acid substitutions are indicated as grey shading.

REFERENCES

Adamson, P. C., Lin, E. Y., Ha, S.-M., and Klausner, J. D. (2021). Using a public database of *Neisseria gonorrhoeae* genomes to detect mutations associated with zoliflodacin resistance. *J. Antimicrob. Chemother.* 76, 2847–2849. doi: 10.1093/jac/dkab262

Alcalá, B., Arreaza, L., Salcedo, C., Antolín, I., Borrell, N., Cacho, J., et al. (2003). Molecular characterization of ciprofloxacin resistance of gonococcal strains in Spain. *Sex. Transm. Dis.* 30, 395–398. doi: 10.1097/00007435-200305000-00004

Allan-Blitz, L.-T., Humphries, R. M., Hemarajata, P., Bhatti, A., Pandori, M. W., Siedner, M. J., et al. (2017). Implementation of a rapid genotypic assay to promote targeted ciprofloxacin therapy of *Neisseria gonorrhoeae* in a large health system. *Clin. Infect. Dis.* 64, 1268–1270. doi: 10.1093/cid/ciw864

Alm, R. A., Lahiri, S. D., Kutschke, A., Otterson, L. G., McLaughlin, R. E., Whiteaker, J. D., et al. (2015). Characterization of the novel DNA gyrase inhibitor AZD0914: low resistance potential and lack of cross-resistance in *Neisseria gonorrhoeae*. *Antimicrob. Agents Chemother.* 59, 1478–1486. doi: 10.1128/AAC.04456-14

Argimón, S., Abudahab, K., Goater, R. J. E., Fedosejev, A., Bhai, J., Glasner, C., et al. (2016). Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* 2:e000093. doi: 10.1099/mgen.0.000093

Belland, R. J., Morrison, S. G., Ison, C., and Huang, W. M. (1994). *Neisseria gonorrhoeae* acquires mutations in analogous regions of gyrA and parC in fluoroquinolone-resistant isolates. *Mol. Microbiol.* 14, 371–380. doi: 10.1111/j.1365-2958.1994.tb01297.x

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., and Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* 13, 42–51. doi: 10.1038/nrmicro3380

Bodoev, I. N., and Il'ina, E. N. (2015). Molecular mechanisms of formation of drug resistance in *Neisseria gonorrhoeae*: history and prospects. *Mol. Genet. Microbiol. Virol.* 30, 132–140. doi: 10.3103/S0891416815030027

Bowler, L. D., Zhang, Q. Y., Riou, J. Y., and Spratt, B. G. (1994). Interspecies recombination between the penA genes of *Neisseria meningitidis* and commensal *Neisseria* species during the emergence of penicillin resistance in *N. meningitidis*: natural events and laboratory simulation. *J. Bacteriol.* 176, 333–337. doi: 10.1128/jb.176.2.333-337.1994

Bradford, P. A., Miller, A. A., O'Donnell, J., and Mueller, J. P. (2020). Zoliflodacin: an oral spiropyrimidinetrione antibiotic for the treatment of *Neisseria gonorrheae*, including multi-drug-resistant isolates. *ACS Infect. Dis.* 6, 1332–1345. doi: 10.1021/acsinfecdis.0c00021

Catlin, B. W. (1960). Interspecific transformation of *Neisseria* by culture slime containing deoxyribonucleate. *Science* 131, 608–610. doi: 10.1126/science.131.3400.608-a

CDC (1985). Tetracycline-resistant *Neisseria gonorrhoeae*—Georgia, Pennsylvania, New Hampshire. *MMWR Morb. Mortal. Wkly Rep.* 34, 563–564.

CDC (1993). CDC sexually transmitted diseases treatment guidelines. Centers for Disease Control and Prevention. *MMWR Recomm. Rep.* 42, 1–102.

Chen, M., Zhang, C., Zhang, X., and Chen, M. (2020). Meningococcal quinolone resistance originated from several commensal *Neisseria* species. *Antimicrob. Agents Chemother.* 64, e01494–e01419. doi: 10.1128/AAC.01494-19

Chesson, H. W., Kirkcaldy, R. D., Gift, T. L., Owusu-Edusei, K. Jr., and Weinstock, H. S. (2014). Ciprofloxacin resistance and gonorrhea incidence rates in 17 cities, United States, 1991-2006. *Emerg. Infect. Dis.* 20, 612–619. doi: 10.3201/eid2004.131288

Davis, R., Markham, A., and Balfour, J. A. (1996). Ciprofloxacin. an updated review of its pharmacology, therapeutic efficacy and tolerability. *Drugs* 51, 1019–1074. doi: 10.2165/00003495-199651060-00010

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., et al. (2020). The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Res.* 48, D606–D612. doi: 10.1093/nar/gkz943

de Korne-Elenbaas, J., Bruisten, S. M., de Vries, H. J. C., and Van Dam, A. P. (2021). Emergence of a *Neisseria gonorrhoeae* clone with reduced cephalosporin susceptibility between 2014 and 2019 in Amsterdam, Netherlands, revealed by genomic population analysis. *J. Antimicrob. Chemother.* 76, 1759–1768. doi: 10.1093/jac/dkab082

Dong, H. V., Pham, L. Q., Nguyen, H. T., Nguyen, M. X. B., Nguyen, T. V., May, F., et al. (2020). Decreased cephalosporin susceptibility of oropharyngeal *Neisseria* species in antibiotic-using men who have sex with men in Hanoi. *Clin. Infect. Dis.* 70, 1169–1175. doi: 10.1093/cid/ciz365

Drlica, K., and Zhao, X. (1997). DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol. Mol. Biol. Rev.* 61, 377–392. doi: 10.1128/mmbr.61.3.377-392.1997

Elhadidy, M., Ali, M. M., El-Shibiny, A., Miller, W. G., Elkhatib, W. F., Botteldoorn, N., et al. (2020). Antimicrobial resistance patterns and molecular resistance markers of campylobacter jejuni isolates from human diarrheal cases. *PLoS One* 15:e0227833. doi: 10.1371/journal.pone.0227833

Fiore, M. A., Raisman, J. C., Wong, N. H., Hudson, A. O., and Wadsworth, C. B. (2020). Exploration of the neisseria resistome reveals resistance mechanisms in commensals that may be acquired by *N. gonorrhoeae* through horizontal gene transfer. *Antibiotics* 9, 1–12. doi: 10.3390/antibiotics9100656

Foerster, S., Drusano, G., Golparian, D., Neely, M., Piddock, L. J. V., Alirol, E., et al. (2019). In vitro antimicrobial combination testing of and evolution of resistance to the first-in-class spiropyrimidinetrione zoliflodacin combined with six therapeutically relevant antimicrobials for *Neisseria gonorrhoeae*. *J. Antimicrob. Chemother.* 74, 3521–3529. doi: 10.1093/jac/dkz376

Foerster, S., Golparian, D., Jacobsson, S., Hathaway, L. J., Low, N., Shafer, W. M., et al. (2015). Genetic resistance determinants, in vitro time-kill curve analysis and pharmacodynamic functions for the novel topoisomerase II inhibitor ETX0914 (AZD0914) in *Neisseria gonorrhoeae*. *Front. Microbiol.* 6:1377. doi: 10.3389/fmicb.2015.01377

Golparian, D., Harris, S. R., Sánchez-Busó, L., Hoffmann, S., Shafer, W. M., Bentley, S. D., et al. (2020). Genomic evolution of *Neisseria gonorrhoeae* since the preantibiotic era (1928-2013): antimicrobial use/misuse selects for resistance and drives evolution. *BMC Genomics* 21:116. doi: 10.1186/s12864-020-6511-6

Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., et al. (2011). Extensive recombination events and horizontal gene transfer shaped the Legionella pneumophila genomes. *BMC Genomics*. doi: 10.1186/1471-2164-12-536

Goytia, M., Thompson, S. T., Jordan, S. V. L., and King, K. A. (2021). Antimicrobial resistance profiles of human commensal *Neisseria* species. *Antibiotics* 10:538. doi: 10.3390/antibiotics10050538

Higashi, D. L., Biais, N., Weyand, N. J., Agellon, A., Sisko, J. L., Brown, L. M., et al. (2011). *N. elongata* produces type IV pili that mediate interspecies gene transfer with *N. gonorrhoeae*. *PLoS One* 6:e21373. doi: 10.1371/journal.pone.0021373

Hong, E., Thulin Hedberg, S., Abad, R., Fazio, C., Enríquez, R., Deghmane, A. E., et al. (2013). Target gene sequencing to define the susceptibility of *Neisseria meningitidis* to ciprofloxacin. *Antimicrob. Agents Chemother.* 57, 1961–1964. doi: 10.1128/AAC.02184-12

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

Irene, G., Marios, G., Fernando, A., Delia, B., Josefina, L., and de la Campa, A. G. (1998). Fluoroquinolone resistance mutations in the parC, parE, and gyrA genes of clinical isolates of viridans group Streptococci. *Antimicrob. Agents Chemother.* 42, 2792–2798. doi: 10.1128/AAC.42.11.2792

Janoir, C., Podglajen, I., Kitzis, M.-D., Poyart, C., and Gutmann, L. (1999). In vitro exchange of fluoroquinolone resistance determinants between *Streptococcus pneumoniae* and Viridans streptococci and genomic organization of the parE-parC region in *S. mitis*. *J. Infect. Dis.* 180, 555–558. doi: 10.1086/314888

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1

José, F. M., Asunción, F., Josefina, L., and De La Campa, A. G. (2000). Horizontal transfer of parC and gyrA in fluoroquinolone-resistant clinical isolates of *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 44, 840–847. doi: 10.1128/AAC.44.4.840-847.2000

Kato, J., Suzuki, H., and Ikeda, H. (1992). Purification and characterization of DNA topoisomerase IV in *Escherichia coli*. *J. Biol. Chem.* 267, 25676–25684. doi: 10.1016/S0021-9258(18)35660-6

Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kenyon, C., Buyze, J., and Wi, T. (2018a). Antimicrobial consumption and susceptibility of *Neisseria gonorrhoeae*: a global ecological analysis. *Front. Med.* 5:329. doi: 10.3389/fmed.2018.00329

Kenyon, C., Laumen, J., and Manoharan-Basil, S. (2021). Choosing new therapies for gonorrhoea: we need to consider the impact on the pan-neisseria genome. a viewpoint. *Antibiotics* 10:515. doi: 10.3390/antibiotics10050515

Kenyon, C. R., Schwartz, I. S., Chris, R., and ISS, K. (2018b). Effects of sexual network connectivity and antimicrobial drug use on antimicrobial resistance in *Neisseria gonorrhoeae*. *Emerg. Infect. Dis.* 24, 1195–1203. doi: 10.3201/eid2407.172104

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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

Lemey, P., Salemi, M., and Vandamme, A. (eds.) (2009). The Phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. 142–160.

Lesher, G. Y., Froelich, E. J., Gruett, M. D., Bailey, J. H., and Brundage, R. P. (1962). 1,8-naphthyridine derivatives. A new class of chemotherapeutic agents. *J. Med. Pharm. Chem.* 91, 1063–1065. doi: 10.1021/jm01240a021

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Ligon, B. L. (2005). Albert Ludwig Sigesmund Neisser: discoverer of the cause of gonorrhea. *Semin. Pediatr. Infect. Dis.* 16, 336–341. doi: 10.1053/j.spid.2005.07.001

Lindbäck, E., Rahman, M., Jalal, S., and Wretlind, B. (2002). Mutations in gyrA, gyrB, parC, and parE in quinolone-resistant strains of *Neisseria gonorrhoeae*. *APMIS* 110, 651–657. doi: 10.1034/j.1600-0463.2002.1100909.x

Manoharan-Basil, S. S., Laumen, J. G. E., Van Dijck, C., De Block, T., De Baetselier, I., Kenyon, C., et al. (2021). Evidence of horizontal gene transfer of 50S ribosomal genes rplB, rplD, and rplY in *Neisseria gonorrhoeae*. *Front. Microbiol.* 12:683901. doi: 10.3389/fmicb.2021.683901

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

Osnes, M, Didelot, X, de Korne-Elenbaas, J, Alfsnes, K., Brynildsrud, O., Syversen, G., et al. (2020). The sudden emergence of a *Neisseria gonorrhoeae* strain with reduced susceptibility to extended-spectrum cephalosporins, Norway. bioRxiv [Preprint]. doi: 10.1101/2020.02.07.935825

Osnes, M. N., van Dorp, L., Brynildsrud, O. B., Alfsnes, K., Schneiders, T., Templeton, K. E., et al. (2021). Antibiotic treatment regimes as a driver of the global population dynamics of a major gonorrhea lineage. *Mol. Biol. Evol.* 38, 1249–1261. doi: 10.1093/molbev/msaa282

Piddock, L. J. (1999). Mechanisms of fluoroquinolone resistance: an update 1994-1998. *Drugs* 58, 11–18. doi: 10.2165/00003495-199958002-00003

Reece, R. J., and Maxwell, A. (1991). DNA gyrase: structure and function. *Crit. Rev. Biochem. Mol. Biol.* 26, 335–375. doi: 10.3109/10409239109114072

Reimche, J. L., Chivukula, V. L., Schmerer, M. W., Joseph, S. J., Pham, C. D., Schlanger, K., et al. (2021). Genomic analysis of the predominant strains and antimicrobial resistance determinants Within 1479 *Neisseria gonorrhoeae* isolates from the US gonococcal isolate surveillance project in 2018. *Sex. Transm. Dis.* 48, S78–S87. doi: 10.1097/OLQ.0000000000001471

Rouquette-Loughlin, C., Dunham, S. A., Kuhn, M., Balthazar, J. T., and Shafer, W. M. (2003). The NorM efflux pump of *Neisseria gonorrhoeae* and *Neisseria meningitidis* recognizes antimicrobial cationic compounds. *J. Bacteriol.* 185, 1101–1106. doi: 10.1128/JB.185.3.1101-1106.2003

Sánchez-Busó, L., Yeats, C. A., Taylor, B., Goater, R. J., Underwood, A., Abudahab, K., et al. (2021). A community-driven resource for genomic epidemiology and antimicrobial resistance prediction of *Neisseria gonorrhoeae* at Pathogenwatch. *Genome Med.* 13:61. doi: 10.1186/s13073-021-00858-2

Salminen, M. O., Carr, J. K., Burke, D. S., and Mccutchan, F. E. (1995). Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Res. Hum. Retroviruses*. doi: 10.1089/aid.1995.11.1423

Shafer, W. M. (2018). Mosaic drug efflux gene sequences from commensal neisseria can lead to low-level azithromycin resistance expressed by *Neisseria gonorrhoeae* clinical isolates. *MBio* 9, e01747–e01718. doi: 10.1128/mBio.01747-18

Shen, Y., and Chen, M. (2020). Prevalence, sequence type, and quinolone resistance of *Neisseria lactamica* carried in children younger than 15 years in Shanghai, China. *J. Inf. Secur.* 80, 61–68. doi: 10.1016/j.jinf.2019.08.020

Shultz, T. R., Tapsall, J. W., and White, P. A. (2001). Correlation of in vitro susceptibilities to newer quinolones of naturally occurring quinolone-resistant *Neisseria gonorrhoeae* strains with changes in GyrA and ParC. *Antimicrob. Agents Chemother.* 45, 734–738. doi: 10.1128/AAC.45.3.734-738.2001

Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* 4:e000166. doi: 10.1099/mgen.0.000166

Tacconelli, E, and Magrini, N. (2017). Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. Organ Mund la Salud.

Tanaka, M., Nakayama, H., Haraoka, M., and Saika, T. (2000). Antimicrobial resistance of *Neisseria gonorrhoeae* and high prevalence of ciprofloxacin-resistant isolates in Japan, 1993 to 1998. *J. Clin. Microbiol.* 38, 521–525. doi: 10.1128/JCM.38.2.521-525.2000

Tapsall, J. W., Ndowa, F., Lewis, D. A., and Unemo, M. (2009). Meeting the public health challenge of multidrug- and extensively drug-resistant *Neisseria gonorrhoeae*. *Expert Rev. Anti-Infect. Ther.* 7, 821–834. doi: 10.1586/eri.09.63

Taylor, S. N., Marrazzo, J., Batteiger, B. E., Hook, E. W. III, Seña, A. C., Long, J., et al. (2018). Single-dose zoliflodacin (ETX0914) for treatment of urogenital gonorrhea. *N. Engl. J. Med.* 379, 1835–1845. doi: 10.1056/NEJMoa1706988

Trees, D. L., Sandul, A. L., Peto-Mesola, V., Aplasca, M. R., Bun Leng, H., Whittington, W. L., et al. (1999). Alterations within the quinolone resistance-determining regions of GyrA and ParC of *Neisseria gonorrhoeae* isolated in the Far East and the United States. *Int. J. Antimicrob. Agents* 12, 325–332. doi: 10.1016/s0924-8579(99)00081-3

Unemo, M., Ahlstrand, J., Sánchez-Busó, L., Day, M., Aanensen, D., Golparian, D., et al. (2021). High susceptibility to zoliflodacin and conserved target (GyrB) for zoliflodacin among 1209 consecutive clinical *Neisseria gonorrhoeae* isolates from 25 European countries, 2018. *J. Antimicrob. Chemother.* 76, 1221–1228. doi: 10.1093/jac/dkab024

Unemo, M, Ross, J, Serwin, A, Gomberg, M, Cusini, M, and Jensen, J.S. (2020). 2020 European guideline for the diagnosis and treatment of gonorrhoea in adults. *Int. J. STD AIDS* doi:10.1177/0956462420949126 [Epub ahead of print].

Unemo, M., and Shafer, W. M. (2011). Antibiotic resistance in *Neisseria gonorrhoeae*: origin, evolution, and lessons learned for the future. *Ann. N. Y. Acad. Sci.* 1230, E19–E28. doi: 10.1111/j.1749-6632.2011.06215.x

Wadsworth, C. B., Arnold, B. J., Sater, M. R. A., and Grad, Y. H. (2018). Azithromycin resistance through interspecific acquisition of an epistasis-dependent efflux pump component and transcriptional regulator in *Neisseria gonorrhoeae*. *MBio* 9, e01419–e01418. doi: 10.1128/mBio.01419-18

Willcox, R. R. (1970). A survey of problems in the antibiotic treatment of gonorrhoea. With special reference to South-East Asia. *Br. J. Vener. Dis.* 46, 217–242. doi: 10.1136/sti.46.3.217

Wittmann, H. G., Stöffler, G., Apirion, D., Rosen, L., Tanaka, K., Tamaki, M., et al. (1973). Biochemical and genetic studies on two different types of erythromycin resistant mutants of *Escherichia coli* with altered ribosomal proteins. *Mol. Gen. Genet.* 127, 175–189. doi: 10.1007/BF00333665

Yahara, K., Ma, K. C., Mortimer, T. D., Shimuta, K., Nakayama, S. I., Hirabayashi, A., et al. (2021). Emergence and evolution of antimicrobial resistance genes and mutations in *Neisseria gonorrhoeae*. *Genome Med.* 13:51. doi: 10.1186/s13073-021-00860-8

Yahara, K, Ma, KC, Mortimer, TD, Shimuta, K., Nakayama, S.I., Hirabayashi, A., et al. (2020). Emergence and evolution of antimicrobial resistance genes and mutations in Neisseria gonorrhoeae. bioRxiv [Preprint]. doi: 10.1101/2020.10.26.354993

Yahara, K., Nakayama, S. I., Shimuta, K., Lee, K. I., Morita, M., Kawahata, T., et al. (2018). Genomic surveillance of neisseria gonorrhoeae to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb. Genomics* 4:e000205. doi: 10.1099/mgen.0.000205

Yang, Y., Liao, M., Gu, W.-M., Bell, K., Wu, L., Eng, N. F., et al. (2006). Antimicrobial susceptibility and molecular determinants of quinolone resistance in *Neisseria gonorrhoeae* isolates from Shanghai. *J. Antimicrob. Chemother.* 58, 868–872. doi: 10.1093/jac/dkl301

Zechiedrich, E. L., and Cozzarelli, N. R. (1995). Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in *Escherichia coli*. *Genes Dev.* 9, 2859–2869. doi: 10.1101/gad.9.22.2859

Zechiedrich, E. L., Khodursky, A. B., and Cozzarelli, N. R. (1997). Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*. *Genes Dev.* 11, 2580–2592. doi: 10.1101/gad.11.19.2580

Zhou, Z., Alikhan, N. F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). Grapetree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fmicb.2022.858857

The HipAB Toxin–Antitoxin System Stabilizes a Composite Genomic Island in *Shewanella putrefaciens* CN-32

Yi Zhao[1], Weiquan Wang[2,3,4], Jianyun Yao[2,3], Xiaoxue Wang[2,3,4], Dong Liu[1] and Pengxia Wang[2,3,4]**

[1]College of Life Sciences, Hebei Normal University, Shijiazhuang, China, [2]Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Key Laboratory of Marine Materia Medica, Innovation Academy of South China Sea Ecology and Environmental Engineering, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China, [3]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China, [4]University of Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:
Daniel Yero,
Universidad Autónoma de Barcelona, Spain

Reviewed by:
Alberto J. Martín-Rodríguez,
Karolinska Institutet (KI), Sweden
Frédéric Goormaghtigh,
University of Basel, Switzerland

***Correspondence:**
Dong Liu
pqw1234@163.com
orcid.org/0000-0003-1546-8215
Pengxia Wang
wangpengxia@scsio.ac.cn
orcid.org/0000-0002-1267-6118

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

Received: 20 January 2022
Accepted: 24 February 2022
Published: 21 March 2022

Citation:
Zhao Y, Wang W, Yao J, Wang X, Liu D and Wang P (2022) The HipAB Toxin–Antitoxin System Stabilizes a Composite Genomic Island in Shewanella putrefaciens CN-32. Front. Microbiol. 13:858857. doi: 10.3389/fmicb.2022.858857

Composite genomic islands (GIs) are useful models for studying GI evolution if they can revert into the previous components. In this study, CGI48—a 48,135-bp native composite GI that carries GI21, whose homologies specifically integrated in the conserved *yicC* gene—were identified in *Shewanella putrefaciens* CN-32. CGI48 was integrated into the $tRNA^{Trp}$ gene, which is a conserved gene locus for the integration of genomic islands in *Shewanella*. Upon expressing integrase and excisionase, CGI48 and GI21 are excised from chromosomes *via* site-specific recombination. The shorter attachment sites of GI21 facilitated the capture of GI21 into CGI48. Moreover, GI21 encodes a functional HipAB toxin–antitoxin system, thus contributing to the maintenance of CGI48 in the host bacteria. This study provides new insights into GI evolution by performing the excision process of the inserting GI and improves our understanding of the maintenance mechanisms of composite GI.

Keywords: ***Shewanella putrefaciens*, mobile genetic element, stability, genomic island, toxin–antitoxin**

INTRODUCTION

Genomic islands (GIs) are discrete DNA segments acquired by horizontal transfer, and they always differ among closely related strains. GIs vary in size from a few to several kilobase pairs and have a mosaic structure that evolves by gene acquisition and loss (Bellanger et al., 2014). Horizontal transfer of GIs can be advantageous for the host, influencing traits, such as pathogenicity, symbiosis, metabolism, phage resistance, and fitness (Dobrindt et al., 2004; Bellanger et al., 2014). Therefore, an understanding of GI evolution is critical for understanding the acquisition of these important adaptive traits.

Composite GI formation is a special type of GI evolution in which one mobile genetic element (MGE) is inserted within another or into the attachment sites of a resident GI (tandem accretion; Bellanger et al., 2014). Many composite GIs have been found through genome comparison (Bellanger et al., 2014), such as the SGI1 variant SGI1-B2 from *Proteus mirabilis*

(Lei et al., 2015), ICE*St1* and CIME302 elements of *Streptococcus thermophilus* (Burrus et al., 2000), and ICE*6013* from *Staphylococcus aureus* (Smyth and Robinson, 2009). The native composite GIs have likely undergone some complicated recombination events; therefore, it is difficult to reconstruct their precise evolutionary history. To date, the formation processes of a few native composite GIs have been determined, such as the tripartite integrative and conjugative element (ICE) assembled through recombination from two GIs with integrases and one ICE without an integrase in *Mesorhizobium ciceri* (Haskett et al., 2016), the tandem structure of GI_{prfC} inserting in the integration site for SXT/R391 ICEs in *Pseudoalteromonas* sp. SCSIO 11900 in our previous study (Wang et al., 2017). Native composite islands that can replicate their evolutionary processes under laboratory conditions would be especially useful for improving our understanding of GI evolution. Interestingly, how composite GIs maintain structural stability should also be explored.

Toxin–antitoxin (TA) systems were originally discovered on conjugative plasmids and participated in their stable maintenance in host bacteria (Ogura and Hiraga, 1983; Roberts et al., 1994). The TA system consists of two neighboring genes, encoding a stable toxin killing the cell or inhibiting cell growth and an unstable antitoxin that masks its toxicity (Wang et al., 2021). A proposed mechanism post-segregationally killing (PSK) was established based on the differential stability of the antitoxin and toxin components. In PSK, plasmid-loss cells do not survive, so the plasmid is maintained in the population (Jurenas et al., 2022). Currently, TA systems have also been found to be ubiquitous in bacterial chromosomes and have been suggested to contribute to the maintenance of integrative MGEs. For example, the MosAT system promotes the maintenance of SXT family ICE carried by some *Vibrio cholerae* strains (Wozniak and Waldor, 2009); the $ParE_{SO}$/$CopA_{SO}$ system stabilizes prophage CP4So in *Shewanella oneidensis* (Yao et al., 2018). Whether the TA system also participates in the maintenance of composite GI is unknown. In this study, a new composite island CGI48 was identified and characterized from *Shewanella putrefaciens* CN32 using genome comparison and excision assay. It evolved by inserting a 21-kb genomic island GI21 into the internal region of CGI48. We further show that GI21 carries a functional HipAB toxin–antitoxin system and contributes to the maintenance of CGI48 in the bacterial host.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

The bacterial strains and plasmids used in this study are listed in **Table 1**. *Shewanella* was grown in LB medium at 30°C. *Escherichia coli* WM3064 was grown in LB medium containing 0.3 mM 2,6-diamino-pimelic acid (DAP) at 37°C. Chloramphenicol (Cm; 30 μg ml^{-1}), kanamycin (50 μg ml^{-1}), and ampicillin (100 μg ml^{-1}) were used in *E. coli*, and chloramphenicol (10 μg ml^{-1}) was used in *Shewanella*. Isopropyl-*β*-D-thiogalactopyranoside (IPTG) was used as an inducer.

TABLE 1 | Strains and plasmids used in this study.

Strains/plasmids	Description[a]	Reference
Shewanella putrefaciens strains		
CN32	*Shewanella putrefaciens* CN32 wild type	Lab stock
Δ*hipAB*	Deletion of *hipAB* genes in CN32	This study
ΔGI21	Deletion of GI21 in CN32	This study
ΔCGI48	Deletion of CGI48 in CN32	This study
CN32 P_{int}::*lacZ*	Integration of plasmid pHGI01 in *int* promoter to monitor the CGI48 and GI21 loss in CN32 wild type	This study
Δ*hipAB* P_{int}::*lacZ*	Integration of plasmid pHGI01 in *int* promoter to monitor the CGI48 and GI21 loss in strain Δ*hipAB*	This study
W3-18-1	*Shewanella putrefaciens* W3-18-1 wild type	Caro-Quintero et al., 2011
ANA3	*Shewanella* sp. ANA-3 wild type	Lab stock
Escherichia coli strains		
WM3064	RP4(tra) in chromosome, DAP-, 37°C	Dehio and Meyer, 1997
K-12 BW25113	lacI^q rrnB$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR*514 Δara*BAD*$_{AH33}$ Δrha*BAD*$_{LD78}$	Baba et al., 2006
Plasmids		
pCA24N	Cm^R; lacI^q, IPTG inducible expression plasmid in *E. coli*	Kitagawa et al., 2005
pHipA	Cm^R; lacI^q, $P_{T5\text{-}lac}$::*hipA*	This study
pHipB	Cm^R; lacI^q, $P_{T5\text{-}lac}$::*hipB*	This study
pHipAB	Cm^R; lacI^q, $P_{T5\text{-}lac}$::*hipA-hipB*	This study
pHGECm	Cm^R; Kan^R; IPTG inducible expression plasmid	Wang et al., 2017
pMD19-T	Amp^R, *E. coli* cloning vector	Invitrogen
pMD19-T-*hipAB*	Amp^R, expressing *hipAB* with its native promoter	This study
$pXis_{21}$	Cm^R, expression plasmid for Xis_{21} from GI21	This study
$pXis_{PO1}$	Cm^R, expression plasmid for Xis_{PO1} from GI*Spu*PO1	This study
$pXis_{ANA3}$	Cm^R, expression plasmid for Xis_{ANA3} from GI*Ssp*ANA3	This study
$pInt_{48}$	Cm^R, expression plasmid for Int_{48}	This study
pHGI01	Kan^R, Integrative *lacZ* reporter plasmid	Fu et al., 2014
pInt2894	Cm^R, expression plasmid for Sputcn32_2894	This study
pHGI01-P_{int}	pHGI01 containing 213 bp upstream of *int* (Sputcn32_2900)	This study
pHGR01	Kan^R, replicative *lacZ* reporter plasmid	Fu et al., 2014
pHGR01-P_{hipA}	Fuse *hipAB* promoter from CN32 with *lacZ* in pHGR01	This study
pK18*mobsacB*-Cm	Km^R, Cm^R, *SacB*, and suicide plasmid used for gene knockout	Wang et al., 2015
pK18Cm-*hipAB*	pK18*mobsacB*-Cm containing the homologous arms of *hipAB*	This study

[a]*Cm^R, chloramphenicol resistance; Kan^R, kanamycin resistance; and Amp^R, ampicillin resistance.*

Construction of Plasmids

The primers used in this study are listed in **Table 2**. The encoding regions of *xis*$_{21}$, *xis*$_{PO1}$, *xis*$_{ANA3}$, *int*$_{48}$, and *int2894* were amplified from the original bacterial host and cloned into the *Eco*RI and *Bam*HI sites of pHGECm using T4 ligase, generating $pXis_{21}$, $pXis_{PO1}$, $pXis_{ANA3}$, and $pInt_{48}$. The encoding regions of *hipA*, *hipB*, and *hipAB* were amplified

from CN32 and inserted into the *Sal*I and *Pst*I sits of pCA24N, generating pHipA, pHipB, and pHipAB. The promoter and encoding region of *hipAB* was amplified from CN32 and inserted into pMD19-T, generating pMD19-T-*hipAB*. To construct the *lacZ* reporter plasmid pHGI01-P_{int}, the reporter region of the integrase gene *Sputcn32_2900* was amplified with the primer pair pHGI01-P_{int}-F/−R and fused with the *lacZ* gene in pHGI01. Then, the integrative plasmid pHGI01-P_{int} was transferred into CN32 and Δ*hipAB* by conjugation and integrated into the promoter region of *Sputcn32_2900*, generating CN32 P_{int}::*lacZ* and Δ*hipAB* P_{int}::*lacZ*. The primer sets mob-F/int-R and Int-F/lacZ-R were used to confirm the construct. To construct pHGR01-P_{hipA}, the promoter of *hipAB* was amplified with primers pHGR01-P_{hipA}-F/−R from CN32, and inserted into the promoterless-*lacZ* reporter plasmid pHGR01.

Construction of *hipAB* Deletion Mutant in CN32

The deletion mutant Δ*hipAB* was constructed based on pK18*mobsacB*-Cm as described previously (Wang et al., 2015). Briefly, the upstream and downstream regions of *hipAB* were amplified from CN32 using the primers listed in **Table 2** and inserted into pK18*mobsacB*-Cm using T4 ligase, producing pK18Cm-*hipAB*. Then, pK18Cm-*hipAB* was introduced into CN32 by conjugation. After mating, cells were spread on LB plates containing Cm to screen the single crossover mutant in which pK18Cm-*hipAB* had integrated into the CN32 genome. The mutant was then grown on LB medium without antibiotics for 8 h. To select mutants in which the second recombination had occurred, the culture was diluted, spread on LB medium containing 10% sucrose, and grown at 30°C for 24–36 h. Single colonies were transferred onto LB- and LB-containing Cm plates simultaneously, and colonies sensitive to Cm were collected and confirmed by PCR followed by DNA sequencing.

Conjugation Assays

The plasmids in this study were transferred from *E. coli* WM3064 into *Shewanella* strains by conjugation assays as described previously (Wang et al., 2015). Briefly, equal amounts of donor and recipient cells were mixed and dropped onto LB medium containing DAP. The plates were incubated at 30°C for 6–8 h, and cells were collected from the lawn and streaked on LB medium with antibiotics to select for transconjugants.

Reporter Activity Assay

Specific *β*-galactosidase activity was determined by monitoring the absorbance at 420 nm using the Miller assay (Miller, 1972). To determine the promoter activity of *hipAB* under overexpression of HipB and HipB-HipA, plasmids pHipB or pHipAB were transformed into the *E. coli* host carrying the reporter plasmid pHGR01-P_{hipA}. Overnight cultures were diluted 1:100 in LB with Kan and Cm and induced with 0.1 mM IPTG at an OD_{600} of 1.0. After induction for 2 h, cells were collected to determine the *β*-galactosidase activity.

Quantification of the Excision Rate of GI21 and CGI48

For GI21, GI*Ssp*ANA3, GI*Spu*PO1, and CGI48, *attB/gyrB* indicated the excision rate of the target GIs after excision. We conducted real-time quantitative PCR (qPCR) assays to quantify the *attB* of these GIs as previously reported methods (Burrus and Waldor, 2003; Wang et al., 2017). The primers used for the qPCR assays are listed in **Table 2**, and chromosomal *gyrB* was used as the reference gene. To test the regulation of Xis and Int on the excision of GI21, GI*Ssp*ANA3, GI*Spu*PO1, and CGI48, $pXis_{21}$-, $pXis_{PO1}$-, $pXis_{ANA3}$-, and $pInt_{48}$-containing strains were induced with 1.0 mM IPTG for 6 h at an OD_{600} of 0.8–1.

Calculation of % CGI48- and GI21-Free Cells

Both CGI48 and GI21 are non-replicable, and loss of CGI48 and GI21 only occurs after their excision. Therefore, to visualize the loss of CGI48 and GI21, the wild-type and Δ*hipAB* strains carrying $pXis_{21}$ or $pInt_{48}$ were induced with 1 mM IPTG for 6 h to overproduce Xis_{21} (to induce GI21 excision) or Int48 (to induce CGI48 excision). Then, the cells were plated on LB plates containing X-gal to calculate the numbers of white colonies (losing CGI48 or GI21) plus blue colonies, and the white colonies were also confirmed by PCR assay.

Plasmid Stability Assay

The contribution of HipA/HipB TA system to plasmid stability was tested as described previously (Yao et al., 2015). Overnight cultures of *E. coli* BW25113 containing plasmid pHipAB or empty vector pCA24N were grown in LB medium with Cm. Then, the preculture was used to inoculate 3 ml LB without antibiotics. Every 12 h of growth, bacterial suspensions were diluted 1,000-fold in 3 ml fresh LB medium. The cultures were serially diluted in 10-fold dilution steps from 0 to 108 h, and 10 μl was dropped on LB plates with or without Cm. The colony-forming unit (CFU) assay was conducted every 12 h for 108 h, and the number of CFUs was determined. Each experiment was performed in triplicate with two independent cultures.

RESULTS

CGI48 Is a Composite Island Containing GI21

Comparing the genome sequence of *S. putrefaciens* CN32 with the related strain *S. putrefaciens* W3-18-1, a large region within 3,340,000–3,400,000 of CN32 was absent in the same gene locus (1,160,000–1,170,000) of W3-18-1 (**Figure 1A**), suggesting that this region was acquired horizontally. Moreover,

TABLE 2 | Primers used in this study.

Primers	Sequence (5'-3')	Purpose
Plasmid construction		
Int48-F	CCGGAATTCATGGGTAGTATTAACTCTCG	$pInt_{48}$
Int48-R	CGCGGATCCTTATCCTCTTAGTTTTTGGTTC	
Xis21-F	CCGGAATTCATGAACCCATCAAATCACG	$pXis_{21}$
Xis21-R	CGCGGATCCTAATTGATACTTTCGCGG	
Int2894-F	CCGGAATTCTTGTCTAAGGACTCGACGGAG	pInt2894
Int2894-R	CGCGGATCCTTATTTGTTGTTCATCATCATTATTCC	
Xis_{PO1}-F	CCGGAATTCGTGAACATGAACCCATCAAATC	$pXis_{PO1}$
Xis_{PO1}-R	CGCGGATCCCTAATTGATACTTTCGCGGTTGG	
Xis_{ANA3}-F	CCGGAATTCGTGAGCATGAACTCATTAAATAAAC	$pXis_{ANA3}$
Xis_{ANA3}-R	CGCGGATCCTTAGTACTTCCCATCTTCGACTG	
hipA-SalI-F	ACGCGTCGACGAACAGTTGACCATTCAGGC	pHipA
hipA-PstI-R	TGCACTGCAGTCATACCAATCCCCAACGCG	
hipB-SalI-F	ACGCGTCGACAGTGATAAACAAACGACTAC	pHipB
hipB-PstI-R	TGCACTGCAGTCATAAAAGCCATGTGACAC	
hipA-SalI-F	ACGCGTCGACGAACAGTTGACCATTCAGGC	pHipAB
hipB-PstI-R	TGCACTGCAGTCATAAAAGCCATGTGACAC	
pHGR01-P_{hipA}-F	CCGGAATTCACTGTAGCGCATATTTAATAA	pHGR01-P_{hipA}
pHGR01-P_{hipA}-R	CGCGGATCCgtaatcatggTCATGAAAGCTCCCAAAGACATTATG	
pMD19-T-*hipAB*-F	TCATAAAAGCCATGTGACAC	pMD19-T-*hipAB*
pMD19-T-*hipAB*-R	GTCACCACATTAGTCCCACT	
Construction of Δ*hipAB*		
hipAB-up-F	ACATGCATGCGAGATGAAACGCTTCAACTCG	pK18Cm-*hipAB*
hipAB-up-R	CCGGAATTCCAGTGGATAGCATTGACCAAC	
hipAB-down-F	CCGGAATTCGCTCGTAATCTAACGAGGTAAG	
hipAB-down-R	AGCGTCGACCCAGGTTACTAATTCTAGTCAC	
hipAB-wF	GTTTACATAAACCAGCAGCAC	Confirmation of Δ*hipAB*
hipAB-wR	GTCCATATTACTCACCTTAGC	
Construction of Δ*hipAB* P_{int}::*lacZ* and CN32 P_{int}::*lacZ*		
pHGI01-P_{int}-F	CCGGAATTCAACGTCGAATGACGTTTTTAGCG	pHGI01-P_{int}
pHGI01-P_{int}-R	CGCGGATCCgtaatcatggGTAGTTAAGTCCAAAATGGTGAC	
mob-F	CAGAGCAGGATTCCCGTTGAGCA	Confirmation of Δ*hipAB* P_{int}::*lacZ* and CN32 P_{int}::*lacZ*
LacZ-R	TATTACGCCAGCTGGCGAAAGG	
Int-F	ATGATTAAGTGTCACTTTTCAAGG	
Int-R	CATTTGGCTGCGATTAGCTC	
Primers used in determination of the excision and circled form of CGI48 and GI21		
21F	CCAAAGCGAGGTAAGACGT	ΔGI21
21R	TCGGAGACAGCGATGTATCG	
21cirF	AGTGGGACTAATGTGGTGACTAGAATT	The circled GI21
21cirR	TGCAAGTGCATGGTTTTATGATG	
48F	CCAAGTGAACGTTTATGATCGC	ΔCGI48
48R	GGTGTGTTTTTCATCGTTATGC	
48cirF	CGAGAGTCTATTCGTAGAGAC	The circled CGI48
48cirR	AGAATATGGTCTAACCAAGC	
oF	CCGGAATTCATGATTAAGTGTCACTTTTCAAGG	*cro/cI* gene
oR	CGCGGATCCTTAGTCTGTACCTTGGATTTC	
Primers used in qPCR for CGI48 and GI21 in CN32		
q48F	GGCTCGCATATTTCTGTGCAA	Determine the excision rate of CGI48
q48R	CCTTTGAGAGTGCTTTTAGCATAATG	
q21F	TTGGCGAGTTGCTCGAAATC	Determine the excision rate of GI21
q21R	GGAACTGGGATGTGTTTTATTGC	
q48cF	CGAGAGTCTATTCGTAGAGAC	Determine the circular form of CGI48
q48cR	AGAATATGGTCTAACCAAGC	
q21cF	AGTGGGACTAATGTGGTGACTAGAATT	Determine the circular form of GI21
q21cR	TGCAAGTGCATGGTTTTATGATG	
CN32gyrB-qF	TTCGTACTTTGCTGTTGACCTTCT	Reference gene
CN32gyrB-qR	CTACGGTGCCATCCAATGCT	
Primers used in qPCR for GI*Spu*PO1 in W3-18-1		
GISpuPO1-qF	AGGTCGCCGTCTCGATTTTA	Determine the excision rate of GI*Spu*PO1
GISpuPO1-qR	TGAGTCGGAAACATCATTAGACGTT	

(Continued)

TABLE 2 | Continued

Primers	Sequence (5'-3')	Purpose
W3181gyrB-qF	GCTCAGCCGCCTTTGTTTAA	Reference gene
W3181gyrB-qR	CGGCTCACCCGACATACC	
Primers used in qPCR for GI*Ssp*ANA3 in ANA-3		
GISspANA3-qF	GTCGAGCTCAAAGTACTCATCGAA	Determine the excision rate of GI*Ssp*ANA3
GISspANA3-qR	GCTACAGCAGAAGCTAATCTCATTACTC	
ANA3gyrB-qF	CTGGTGAGCCTGTGCTCGAT	Reference gene
ANA3gyrB-qR	CAAGCGCCGCACCTAACTTA	

Restriction sites included in oligonucleotide sequences are underlined.

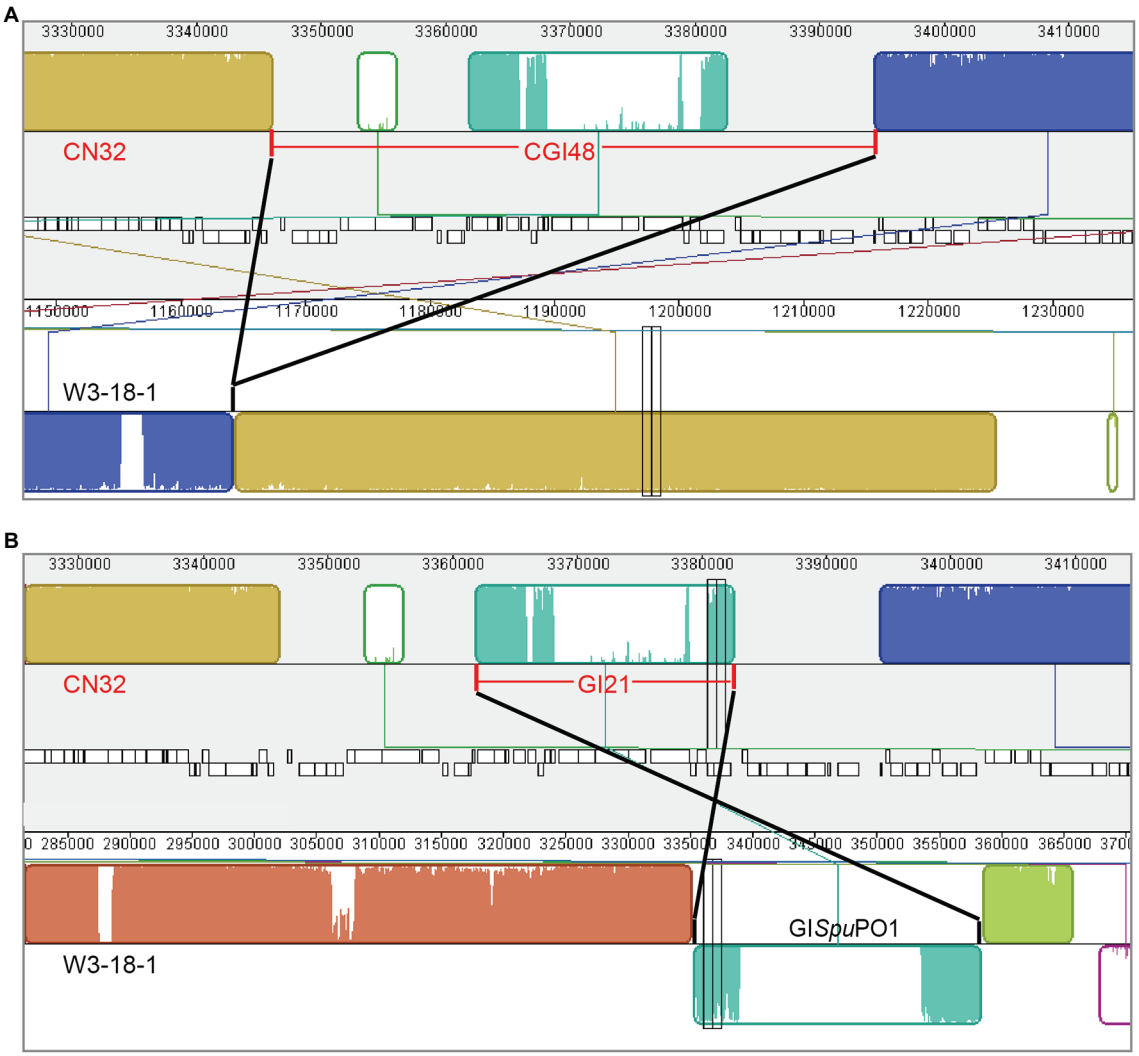


FIGURE 1 | Schematic view of the composite island CGI48 in the CN32 genome. CGI48 **(A)** and its component genomic island GI21 **(B)** were identified by comparing the genome sequence of *S. putrefaciens* CN32 (CP000681) with *S. putrefaciens* W3-18-1 (CP000503) with Mauve.

the internal sequence within 3,360,000–3,380,000 of this large region showed high homology with another region 335,000–360,000 of W3-18-1 (**Figure 1B**). These results suggested that the region within 3,340,000–3,400,000 of CN32 is a putative composite genomic island. It is 48 kb in length; thus, it is designated CGI48 hereafter (**Table 3**). Further analysis showed that region 3,360,000–3,380,000 of CN32 contains a 21 kb genomic island (designated GI21), which

TABLE 3 | Sequence analysis of composite island CGI48.

Gene	Start	End	Strand	Functions
$attL_{48}$	3,346,221	3,346,273	+	Left attachment site of CGI48
Sputcn32_2886	3,346,726	3,347,073	+	Hypothetical protein
Sputcn32_2887	3,347,409	3,347,561	+	Pseudo
Sputcn32_2888	3,348,921	3,347,632	–	Beta-lactamase domain protein
Sputcn32_2889	3,349,853	3,348,921	–	Hypothetical protein
Sputcn32_2890	3,350,524	3,349,859	–	Metallophosphoesterase
Sputcn32_2891	3,351,168	3,350,566	–	Conserved hypothetical protein
Sputcn32_2892	3,351,518	3,352,093	+	Hypothetical protein
Sputcn32_2893	3,352,083	3,354,299	+	Hypothetical protein
Sputcn32_2894	3,354,292	3,357,321	+	Phage integrase
Sputcn32_2895	3,357,533	3,358,846	+	Conserved hypothetical protein
Sputcn32_2896	3,359,598	3,359,224	–	Conserved hypothetical protein
Int_{48}, Sputcn32_2897	3,361,269	3,360,109	–	Phage integrase
Sputcn32_2898	3,361,484	3,361,278	–	Transcription-repair coupling factor (superfamily II helicase)
Sputcn32_2899	3,361,613	3,361,819	+	Predicted transcriptional regulator, Cro/CI family
$attL_{21}$	3,361,829	3,361,837	+	Left attachment site of GI21
Sputcn32_2900[a]	3,362,021	3,363,319	+	Phage Integrase
Sputcn32_2901[a]	3,363,329	3,364,162	+	Hypothetical protein
Xis_{21}, Sputcn32_2902[a]	3,364,278	3,364,487	+	AlpA family phage transcriptional regulator
Sputcn32_2903[a]	3,364,910	3,365,845	+	Hypothetical protein
Sputcn32_2904[a]	3,366,023	3,366,676		Conserved hypothetical protein
Sputcn32_2905[a]	3,367,222	3,366,839	–	Hypothetical protein
Sputcn32_2906[a]	3,367,382	3,367,798	+	Putative DNA-binding protein
Sputcn32_2907[a]	3,367,890	3,368,189	+	Protein of unknown function UPF0150
Sputcn32_2908[a]	3,368,533	3,370,104	+	Type I restriction-modification system, M subunit, N-6 DNA methylase
Sputcn32_2909[a]	3,370,094	3,371,416	+	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)
Sputcn32_2910[a]	3,371,431	3,374,133	+	ATPase associated with various cellular activities, AAA_5″
Sputcn32_2911[a]	3,374,133	3,375,440	+	Conserved hypothetical protein
Sputcn32_2912[a]	3,375,839	3,378,976	+	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)
Sputcn32_2913[a]	3,379,461	3,379,039	–	Transcriptional regulator, XRE family
Sputcn32_2914[a]	3,379,625	3,380,281	+	Conserved hypothetical protein
HipB, Sputcn32_2915[a]	3,380,923	3,380,465	–	Transcriptional regulator, XRE family
HipA, Sputcn32_2916[a]	3,382,266	3,380,920	–	HipA domain protein
$attR_{21}$	3,382,740	3,382,748	+	Right attachment site of GI21
Sputcn32_2917	3,383,263	3,383,625	+	Conserved hypothetical protein
Sputcn32_2918	3,384,571	3,383,654	–	Transposase, IS4 family
Sputcn32_2919	3,385,229	3,384,696	–	Conserved hypothetical protein
Sputcn32_2,920	3,386,826	3,385,240	–	Von Willebrand factor, type A
Sputcn32_2921	3,388,288	3,386,819	–	ATPase associated with various cellular activities, AAA_5
Sputcn32_2922	3,389,920	3,388,445	–	Sigma54 specific transcriptional regulator, Fis family
Sputcn32_2923	3,390,290	3,390,066	–	Hypothetical protein
Sputcn32_2924	3,390,847	3,390,707	–	Pseudo
Sputcn32_2925	3,392,615	3,390,858	–	Methyltransferase type 11
Sputcn32_2926	3,393,959	3,393,004	–	Pseudo
Sputcn32_2927	3,394,306	3,394,382	+	tRNA-Trp
$attR_{48}$	3,394,303	3,394,355		Right attachment site of CGI48

[a]*The genes in GI21.*

exhibits sequence identity with genomic islands integrated in the conserved *yicC* gene, such as GI*Spu*PO1 in *S. putrefaciens* W3-18-1, GI*Ssp*ANA3 in *Shewanella* sp. ANA-3, and GI*Psp*SM9913 in *Pseudoalteromonas* sp. SM9913 (**Figure 2A**). GI21 exhibits 99% sequence identity with the two ends of GI*Spu*PO1 in W3-18-1. The left region of GI21 contains an integrase and an excisionase gene next to the left attachment site ($attL_{21}$), and the right region contains a putative *hipA-hipB* toxin–antitoxin pair next to the right attachment site ($attR_{21}$). The middle region contains 12 genes encoding a restriction–modification system and hypothetical proteins (**Figure 2A**).

Excision of GI followed by formation of circular forms of GI is prequisite for its horizontal transfer. Integrase is essential for GI excision and integration, and some GIs also encode recombination directionality factors (or excisionases Xis) directing the reaction toward excision (Lewis and Hatfull, 2001). We wondered whether GI21 can be excised from the CGI48 genome by recombining the attachment $attL_{21}$ and $attR_{21,}$ and producing $attB_{21}$ and $attP_{21}$ sites (**Figure 2B**). Quantitative PCR (qPCR) was used to quantify the excision rate by measuring the percentage of cells in the culture containing $attB_{21}$, which is only present after GI21 excision. In this assay, the amount of $attB_{21}$ sites is compared to the

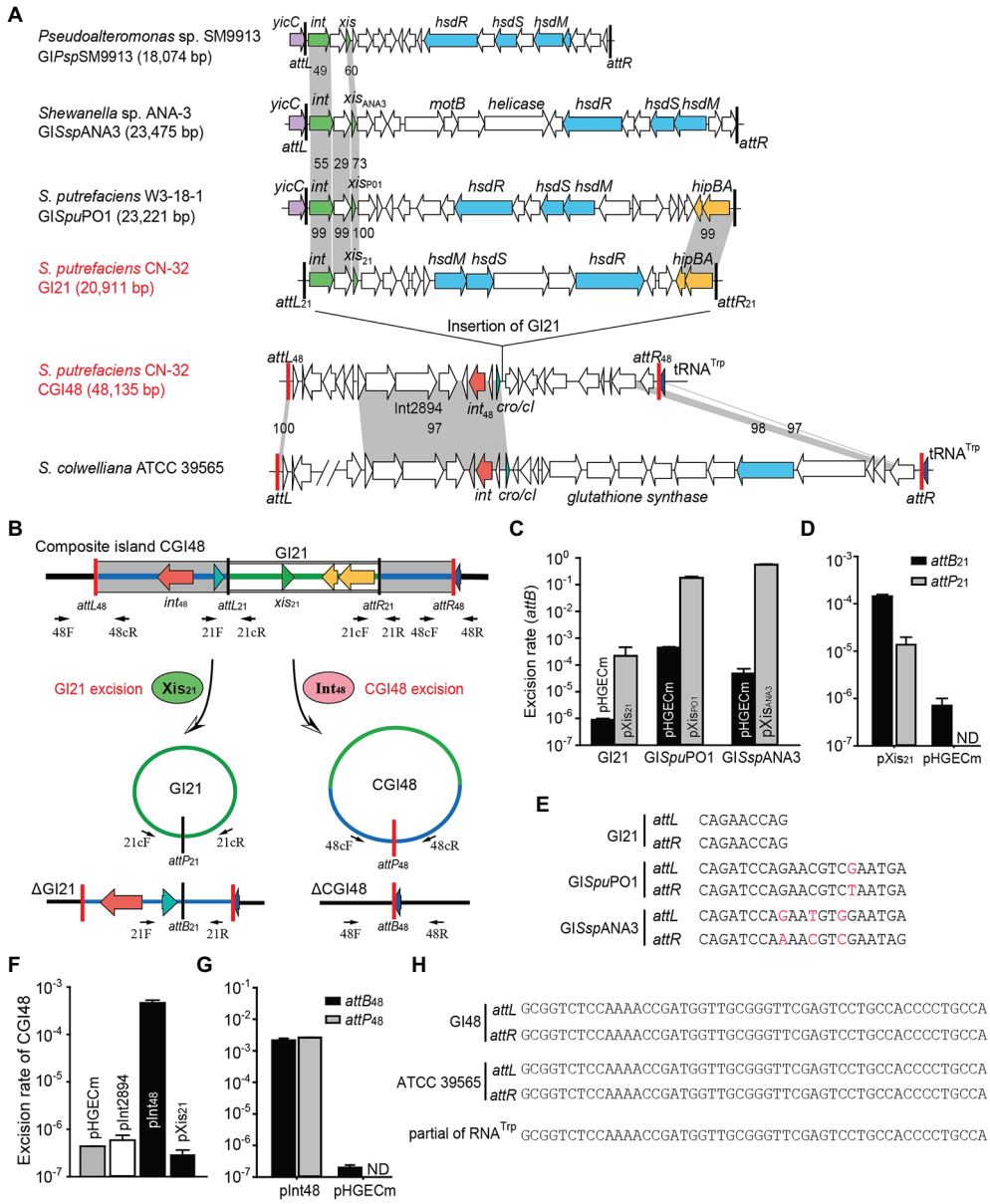


FIGURE 2 | The composite island CGI48 and its component GI21 can be excised from the CN32 genome. **(A)** Sequence analysis of CGI48 with the related genomic islands. Open reading frames with putative functions are shown in different colors. The *attL* and *attR* attachment sites of CGI48 and GI21 are shown in red and black, respectively. The sequence of the GI integrated into $tRNA^{Trp}$ of ATCC 39565 genome was in L876DRAFT_scaffold 00018.18_C (82,217–85,620 bp) to scaffold 00033.33_C (1,478–38,068 bp). **(B)** Schematic of the excision of CGI48 and GI21. **(C)** The excision rate of GI21, GI*Spu*PO1, and GI*Ssp*ANA3 was quantified when cognate excisionase was overexpressed. **(D)** Comparison of the excision rate ($attB_{21}$) and circular form of GI21 ($attP_{21}$) in CN32 when Xis21 was overexpressed. ND indicates not detected. **(E)** Sequence comparison of the attachments of GI21, GI*Spu*PO1, and GI*Ssp*ANA3. **(F)** The excision rate of CGI48 when Int_{48} was overexpressed. **(G)** Comparison of the excision rate ($attB_{48}$) and circular form of CGI48 ($attP_{48}$) in CN32 when Int48 was overexpressed. ND indicates not detected. **(H)** Sequence comparison of the attachment sites of CGI48 and GI in ATCC 39565 compared with the 3′ end of RNA^{Trp} in *Shewanella*.

amount of the reference gene *gyrB*, which is used to quantify the total number of cells in the culture. Excisionase Sputcn32_2902 (Xis_{21}) was induced in strain CN32 with 1 mM IPTG for 6 h. Additionally, the excisionases Xis_{PO1} and Xis_{ANA3} were also overexpressed in W3-18-1 and ANA-3 as a control. The results showed that Xis_{21} mediated GI21 excision, resulting in a 440-fold increase in the excision rate of GI21 and reaching $(3.8 \pm 0.3) \times 10^{-4}$. However, the excision rate of GI*Spu*PO1 and GI*Ssp*ANA3 reached 17.9%–55.6% when Xis_{PO1} and Xis_{ANA3} were overexpressed, which was much higher than that of GI21 (**Figure 2C**). qPCR was also used to quantified the circular form of GI21 by

measuring $attP_{21}$, which is present after GI21 is circularized or replicated after excision. The number of $attP_{21}$ is less than $attB_{21}$, suggesting that GI21 is non-replicable in wild-type CN32 or expressing Xis_{21} (**Figure 2D**). PCR sequencing showed that the attachment sites of GI*Spu*PO1 and GI*Ssp*ANA3 were 21 bp in length, and the attachment sites of GI21 were 9 bp (**Figure 2E**). In CGI48, GI21 was integrated in the untranslated region between Sputcn32_2899 and Sputcn32_2917, which encoded a predicted transcriptional regulator of the Cro/CI family and a conserved hypothetical protein, respectively (**Figure 2A**). The excision and integration of GI21 did not cause any sequence changes in the neighboring genes. The results suggested that GI21 can be excised from CN32 by site-specific recombination of $attL_{21}$ and $attR_{21}$, and the shorter attachment sites may greatly limit the recombination efficiency.

We then evaluated the excision of CGI48 (**Figure 2B**), and the integrase genes *Sputcn32_2894* and *Sputcn32_2897* were cloned into pHGECm for their overexpression. Overproduction of Sputcn32_2897 (named Int_{48}) resulted in a 1.070-fold increase in the excision rate of CGI48 and reached $(4.7 \pm 0.6) \times 10^{-4}$, and Sputcn32_2894 did not affect the excision of CGI48 (**Figure 2F**). Quantification of $attP_{48}$ indicated that CGI48 is non-replicable in wild-type CN32 or expressing Int_{48} (**Figure 2G**). Sequence analysis showed that CGI48 was integrated in the 5′ end of $tRNA^{Trp}$, a conserved integration locus of GIs, such as the GI in *S. colwelliana* ATCC 39565 (**Figure 2A**). PCR sequencing confirmed that CGI48 and GI in *S. colwelliana* ATCC 39565 shared 100% identical attachment sites of 50 bp in length, and the excision and integration did not cause sequence changes in $tRNA^{Trp}$ (**Figure 2H**). Phylogenetic tree analysis of Int21 and Int48 revealed that GI21 homologs are widely distributed in *Shewanella*, *Pseudoalteromonas*, *Halomonas*, and *Vibrio* strains (**Figure 3A**), and CGI48 homologs are widely distributed in *Shewanella*, *Pseudomonas*, *Halomonas*, and *Photobacterium* strains (**Figure 3B**). Collectively, CGI48 and the component GI21 can be excised from the CN32 genome, suggesting that CGI48 is an active composite island in host bacteria.

GI21 Encodes a HipAB Toxin–Antitoxin System

In GI21, two neighboring genes that are only 4 bp apart, *Sputcn32_2916* and *Sputcn32_2915*, were identified as a putative *hipA-hipB* TA pair. In HipA/HipB TA system characterized in *E. coli* K-12, $HipA_{K\text{-}12}$ toxin functions as a serine/threonine protein kinase that inhibits cell growth, and $HipB_{K\text{-}12}$ antitoxin encoded by the gene upstream to *hipA* blocks its effects (Germain et al., 2013). Here, the putative *hipA-hipB* TA pair in GI21 has a genetic architecture reversed to that of *hipB-HipA* in *E. coli* K-12 (**Figure 4A**). *Sputcn32_2916* encodes a HipA domain protein that is 448 aa in length, and it has 40% identity and 6% coverage with $HipA_{K\text{-}12}$. *Sputcn32_2915* encodes a XRE family transcriptional regulator of 152 aa that contains a Helix-turn-helix (HTH) domain in the C-terminal and has 33% amino acid sequence identity and 23% coverage with $HipB_{K\text{-}12}$ (**Figure 4B**). To determine whether *Sputcn32_2916* and *Sputcn32_2915* constitute a functional TA pair, open reading frames of the two genes were cloned into plasmid pCA24N to obtain pHipA and pHipB, respectively. Expression of *hipA* or *hipB* was induced in *E. coli* BW25113 with 0.5 mM IPTG. Cell growth (turbidity) and cell viability (CFU ml^{-1}) were measured for 8 h. Overproducing HipA in BW25113 cells led to growth inhibition (**Figures 4C–E**). To further assess whether HipB can block the toxicity of HipA, we cloned the coding regions of *hipA* and *hipB* into plasmid pCA24N to construct pHipAB. Coexpression of *hipA* and *hipB via* plasmid pHipAB in BW25113 cells showed that HipB could partially neutralize the toxic effect of HipA (**Figures 4C–E**); this may result from the too high load of toxins driven by the strong *lac* promoter on the high copy number plasmid pCA24N. Then, we cloned *hipA-hipB* with its native promoter into pMD19-T to generate pMD19-T-*hipAB*. The strain BW25113/pMD19-T-*hipAB* exhibited similar cell viability with that of BW25113/pMD19-T, suggesting that HipB could fully neutralize the toxic effect of HipA under the native promoter (**Figure 4F**). Taken together, HipA and HipB in GI21 form a TA pair in which HipA is a potent toxin and HipB is the cognate antitoxin.

In $HipA_{K\text{-}12}$/$HipB_{K\text{-}12}$, the antitoxin $HipB_{K\text{-}12}$ or the TA complex bind DNA and autoregulate the transcription of the TA operon (Black et al., 1994). Similar to $HipB_{K\text{-}12}$, HipB in CN32 also contains a HTH domain, thus we wondered whether HipB in GI21 can regulate the *hipA-hipB* operon. Using the plasmid by fusing *lacZ* with the *hipA-hipB* promoter as the reporter, we found that overproduction of HipB exhibited 2.1 ± 0.1-fold decrease in the promoter activity compared to empty vector. Moreover, overproduction of HipA/HipB complex *via* pHipAB showed a 2.9 ± 0.4-fold decrease in the promoter activity (**Figure 4G**). These results suggested that GI21-encoded HipB and the HipA/HipB complex can repress the TA operon.

GI21-Encode HipAB Stabilizes CGI48

To test whether the HipA/HipB TA system affects the excision of CGI48, we deleted the *hipAB* region in CN32. qPCR assays showed no significant difference in the excision rate of CGI48 in the *hipAB* deletion mutant compared to wild-type CN32 (**Figure 5A**). As reported in our previous study, the TA system in prophage CP4So in *S. oneidensis* stabilizes CP4So after its excision (Yao et al., 2018). We wondered whether GI21-encoding HipA/HipB played a role in the maintenance of CGI48 after its excision. A blue–white reporter screening assay was designed to detect the loss of GI21 and CGI48 after their excision. In brief, the *lacZ* gene was fused with the promoter of the integrase gene *Sputcn32_2900* to generate a P_{int}::*lacZ* fusion and cloned into the integrative plasmid pHGI01, generating pHGI01-P_{int}. The constructed plasmid was site specifically integrated into GI21 in CN32. Blue colonies indicated the presence of GI21 in CN32, irrespective of whether it was integrated in the host

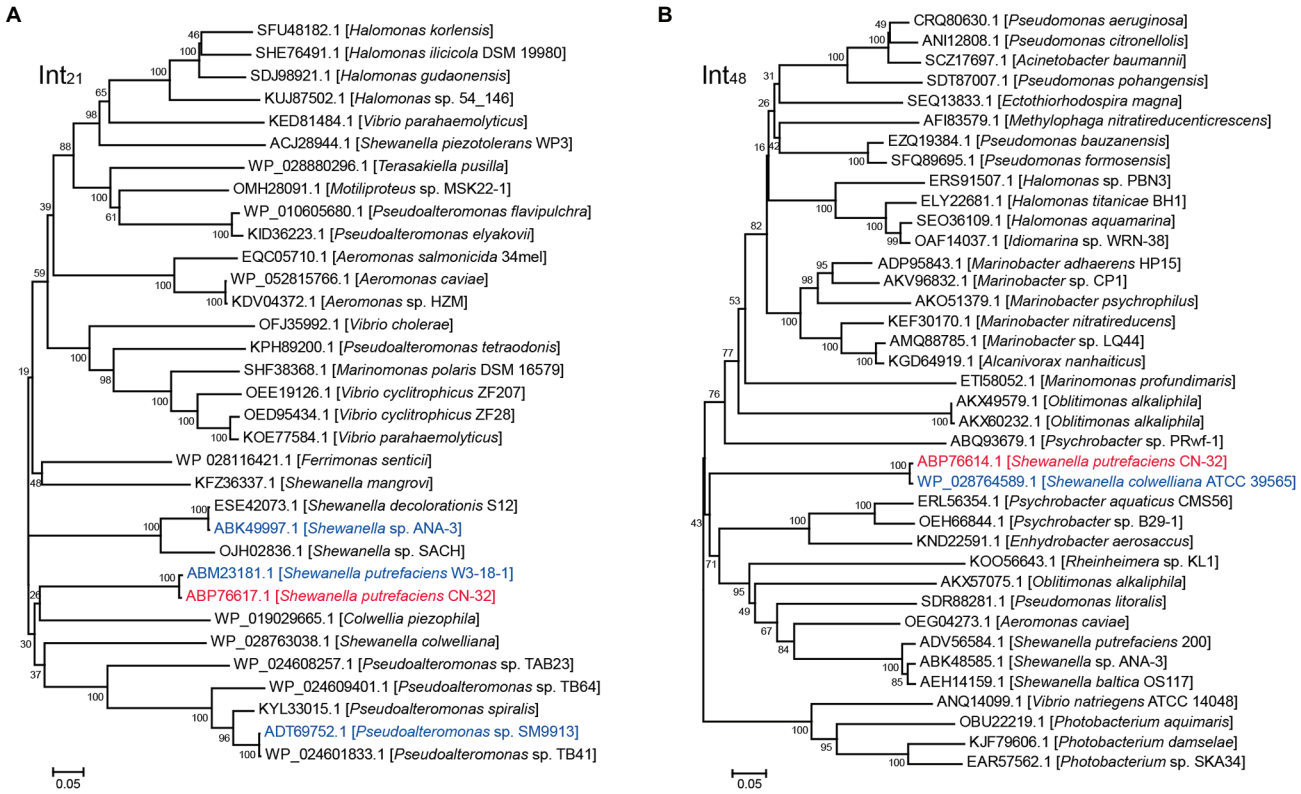


FIGURE 3 | Phylogenetic analysis of Int_{21} and Int_{48} homologs. **(A)** Neighbor-joining phylogenetic tree of 32 homologs of Int_{21}, and **(B)** 37 homologs of Int_{48} based on amino acid (aa) sequences. Proteins with homology to Int_{21} and Int_{48} were selected by BLASTp at the NCBI server. CN32 was indicated in red, and other genomic islands shown in **Figure 2A** were indicated in blue.

chromosome or existed in a circular form after GI21 or CGI48 was excised. White colonies indicated a complete loss of GI21 from CGI48 or a complete loss of CGI48 from the CN32 genome (**Figure 5B**). To activate the excision of GI21 and CGI48, Xis_{21} and Int_{48} were induced with 1 mM IPTG for 6 h, and cells were then plated on X-gal plates to detect GI21- and CGI48-free cells using the reporter plasmid (**Figure 5C**). No loss of GI21 was detected in wild-type CN32, and 0.39% of GI21-free cells were exhibited in the *hipAB* deletion mutant when Xis_{21} was overexpressed. Similarly, no loss of CGI48 was detected in wild-type CN32, and 0.82% of CGI48-free cells was exhibited in the *hipAB* deletion mutant when Int48 was overexpressed (**Figure 5D**). Then, two white colonies (indicated with blue arrows) from the Xis_{21}-induced plates and two (indicated with blue arrows) from the Int_{48}-induced plates were randomly selected to confirm the loss of GI21 and CGI48 (**Figure 5E**) by PCR. In addition, we also test the contribution of GI21-encoded HipA/HipB on plasmid stability. As shown in **Figure 5F**, plasmid pCA24N was completely lost from *E. coli* BW25113 after 72 h, while pHipAB which contains *hipAB* in pCA24N was stably maintained in *E. coli* after 108 h of culturing. Altogether, these results thus demonstrate that HipA/HipB not only stabilizes GI21 and CGI48 but also provides plasmid stabilization.

CONCLUSION AND DISCUSSION

In this study, a new composite island, CGI48, was detected in the genome of *S. putrefaciens* CN-32. CGI48 harbors genes encoding adaptive traits, such as antibiotics and restriction–modification systems. CGI48 evolved by inserting a genomic island, GI21, showing high identity with GIs integrated in the *yicC* locus. Because the conserved *yicC* locus is intact and available in CN32 genome, GI21 might integrated into CN32 accompanied by the composite CGI48. Another possibility is that GI21 is integrated into the secondary attachment site within CGI48 genome after horizontal gene transfer. Many genomic islands preferentially integrated into a primary attachment site in the bacterial genome. Studies on the ICEs, ICE*Bs*1 found that ICE*Bs*1 can also integrate into secondary attachment site, especially when the primary site is absent. However, the excision of ICE*Bs*1 from secondary sites is greatly reduced compared to the primary site, limiting the dissemination of ICE*Bs*1 (Menard and Grossman, 2013). *In vitro* assays showed that the efficiency of integrase-mediated site-specific recombination is related to the length of the attachment site, and the reduction of the core attachment site produced a dramatically decrease in the recombination activity (Ghosh et al., 2003). Thus, we speculated that the shorter attachment sites flanking GI21 may limit its excision

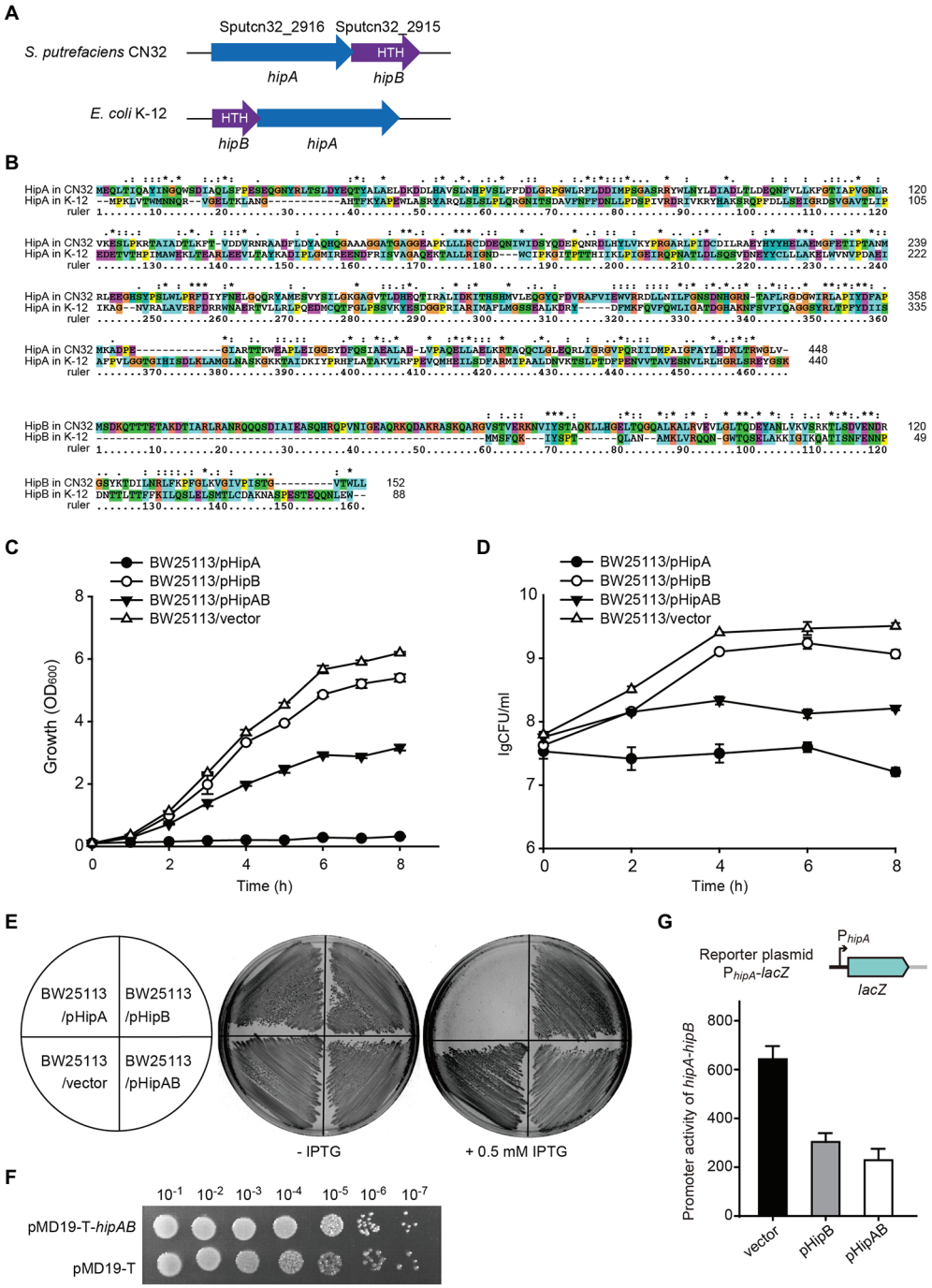


FIGURE 4 | HipA and HipB in GI21 constitute a Toxin–antitoxin (TA) pair. **(A)** Comparison of the *hipA-hipB* operon in GI21 and *hipB-hipA* operon in *E. coli* K-12. **(B)** Sequence alignment was carried out using ClustalX to compare the amino acid sequence identity of HipA/HipB in *S. putrefaciens* CN32 and *E. coli* K-12. Cell growth **(C)** and cell viability **(D)** of cells overexpressing *hipA*, *hipB*, and *hipA-hipB via* pCA24N-based plasmids in *E. coli* BW25113. **(E)** Growth of BW25113 cells overexpressing *hipA*, *hipB*, and *hipA-hipB via* pCA24N-based plasmids on LB plates with and without 0.5 mM isopropyl-*β*-D-thiogalactopyranoside (IPTG). **(F)** CFU of strain BW25113 containing pMD19-T-*hipAB* or empty vector pMD19-T on LB plates with ampicillin. **(G)**The activity of the *hipA-hipB* promoter in GI21 was measured by overexpressing *hipB* or *hipA-hipB*.

and stabilize the composite structure. Some composite GIs are also found to be stabilized by truncated attachment sites or integrases (Bellanger et al., 2014). In this study, we also found that a functional TA system maintain the stability of the composite GI. All these mechanisms explain the complexity and diversity of GIs.

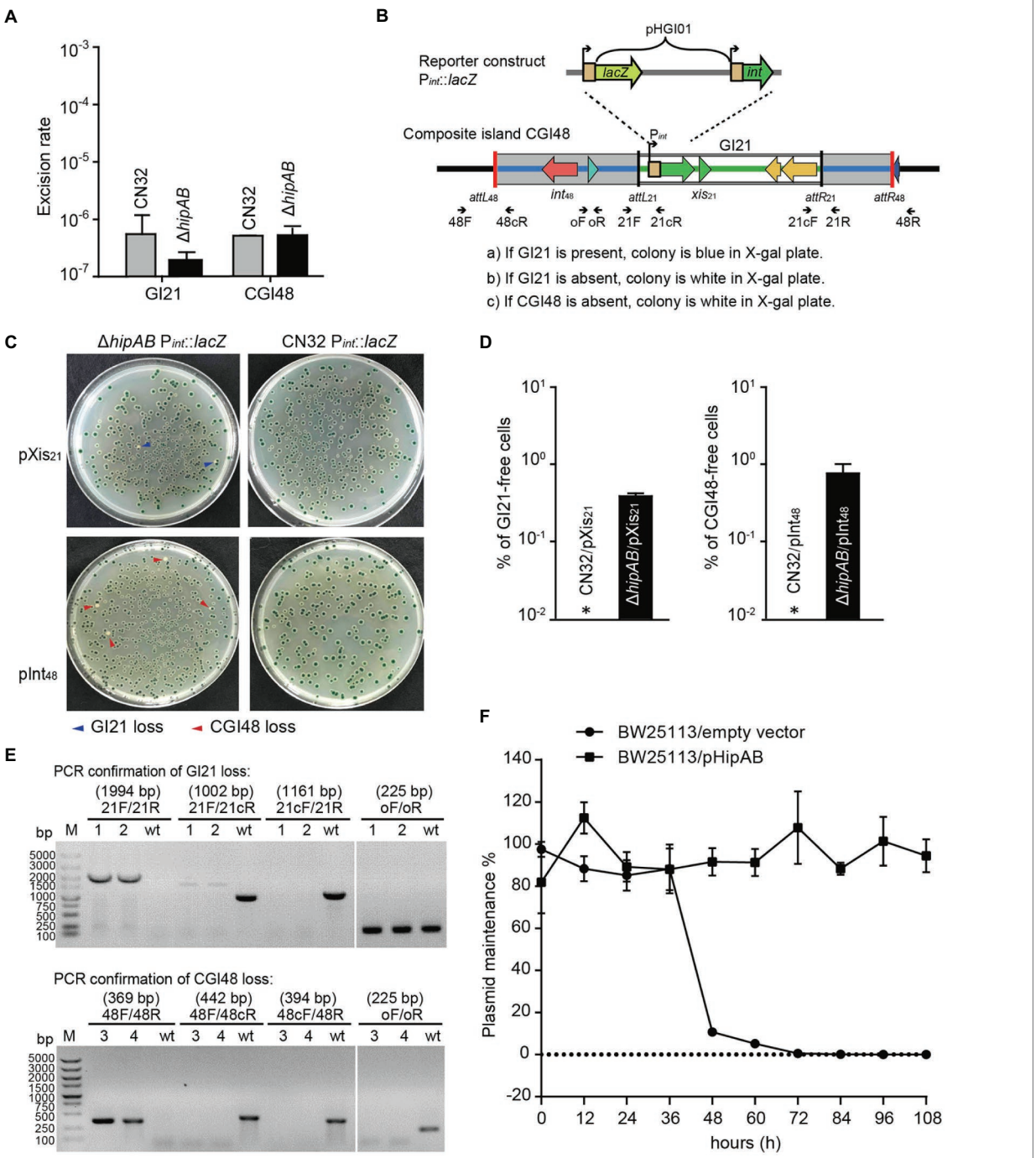


FIGURE 5 | GI21-encoded HipAB promotes the maintenance of CGI48. **(A)** The excision rate of GI21 and CGI48 in the CN32 wild-type and Δ*hipAB* mutant strains. **(B)** Schematic of the *lacZ* reporter constructs in the CN32 wild-type and Δ*hipAB* strains. **(C)** Observation of GI21 loss when Xis_{21} is overexpressed (upper plates) and of CGI48 loss when Int_{48} is overexpressed (lower plates) on X-gal plates using the *lacZ* reporter system. **(D)** % of GI21-free cells (left panel) and % CGI48-free cells (right panel) were quantified by counting five plates, a representative image as shown in **(C)**. Asterisks indicate that the frequency of GI21 and CGI48 loss was below the limit of detection of the assays ($<1 \times 10^{-5}$). **(E)** Confirmation of GI21 (upper panel) and CGI48 (lower panel) loss by PCR using the indicated primers in **(B)**. 1 and 2 indicate the DNA templates extracted from the colonies with blue arrows in **(C)**; 3 and 4 indicate the DNA templates extracted from the colonies with red arrows in **(C)**; wt indicates the DNA templates from wild-type CN32 used as a control. Lane M indicates DNA Marker DL5k. The expected product sizes are indicated at the top of the primer sets. **(F)** GI21-encoded HipAB confers plasmid stability in *E. coli*. *E. coli* BW25113 harboring plasmids pHipAB and empty vector pCA24N were used in this assay. Three independent cultures were conducted, and the data are shown as means ± SDs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/genbank/, CP000503; https://www.ncbi.nlm.nih.gov/genbank/, CP000681.

AUTHOR CONTRIBUTIONS

XW and PW conceptualized and designed the project. YZ, WW, JY, XW, DL, and PW did the investigation, data curation, and analysis. YZ, XW, DL, and PW wrote, reviewed, and edited the original draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Guangdong Major Project of Basic and Applied Basic Research (2019B030302004), the Natural Science Foundation of Guangdong Province (2019A1515011912), the Science and Technology Planning Project of Guangzhou (202002030493), Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City (320LH047), the Youth Innovation Promotion Association CAS (2021345 to PW), the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou; GML2019ZD0407), the Natural Science Foundation of Hebei Province (C2019205044), Research Fund of Hebei Normal University (L2016Z03), and Science and Technology Research Project of Hebei University (ZD2018070).

REFERENCES

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2:2006.0008. doi: 10.1038/msb4100050

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guedon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058

Black, D. S., Irwin, B., and Moyed, H. S. (1994). Autoregulation of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis. *J. Bacteriol.* 176, 4081–4091. doi: 10.1128/jb.176.13.4081-4091.1994

Burrus, V., Roussel, Y., Decaris, B., and Guedon, G. (2000). Characterization of a novel integrative element, ICE*St1*, in the lactic acid bacterium *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* 66, 1749–1753. doi: 10.1128/AEM.66.4.1749-1753.2000

Burrus, V., and Waldor, M. K. (2003). Control of SXT integration and excision. *J. Bacteriol.* 185, 5045–5054. doi: 10.1128/JB.185.17.5045-5054.2003

Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Hofle, M. G., Klappenbach, J., et al. (2011). Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *ISME J.* 5, 131–140. doi: 10.1038/ismej.2010.93

Dehio, C., and Meyer, M. (1997). Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn*5* in *Bartonella henselae* following conjugal plasmid transfer from *Escherichia coli*. *J. Bacteriol.* 179, 538–540. doi: 10.1128/jb.179.2.538-540.1997

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414–424. doi: 10.1038/nrmicro884

Fu, H., Jin, M., Ju, L., Mao, Y., and Gao, H. (2014). Evidence for function overlapping of CymA and the cytochrome bc_1 complex in the *Shewanella oneidensis* nitrate and nitrite respiration. *Environ. Microbiol.* 16, 3181–3195. doi: 10.1111/1462-2920.12457

Germain, E., Castro-Roa, D., Zenkin, N., and Gerdes, K. (2013). Molecular mechanism of bacterial persistence by HipA. *Mol. Cell* 52, 248–254. doi: 10.1016/j.molcel.2013.08.045

Ghosh, P., Kim, A. I., and Hatfull, G. F. (2003). The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of *attP* and *attB*. *Mol. Cell* 12, 1101–1111. doi: 10.1016/s1097-2765(03)00444-1

Haskett, T. L., Terpolilli, J. J., Bekuma, A., O'hara, G. W., Sullivan, J. T., Wang, P., et al. (2016). Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12268–12273. doi: 10.1073/pnas.1613358113

Jurenas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Microbiol.* doi:10.1038/s41579-021-00661-1 [Epub ahead of print].

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., et al. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 12, 291–299. doi: 10.1093/dnares/dsi012

Lei, C. W., Zhang, A. Y., Liu, B. H., Wang, H. N., Yang, L. Q., Guan, Z. B., et al. (2015). Two novel *salmonella* genomic island 1 variants in *Proteus mirabilis* isolates from swine farms in China. *Antimicrob. Agents Chemother.* 59, 4336–4338. doi: 10.1128/AAC.00120-15

Lewis, J. A., and Hatfull, G. F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and cox proteins. *Nucleic Acids Res.* 29, 2205–2216. doi: 10.1093/nar/29.11.2205

Menard, K. L., and Grossman, A. D. (2013). Selective pressures to maintain attachment site specificity of integrative and conjugative elements. *PLoS Genet.* 9:e1003623. doi: 10.1371/journal.pgen.1003623

Miller, J.H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press

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

Roberts, R. C., Strom, A. R., and Helinski, D. R. (1994). The *parDE* operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *J. Mol. Biol.* 237, 35–51. doi: 10.1006/jmbi.1994.1207

Smyth, D. S., and Robinson, D. A. (2009). Integrative and sequence characteristics of a novel genetic element, ICE*6013*, in *Staphylococcus aureus*. *J. Bacteriol.* 191, 5964–5975. doi: 10.1128/JB.00352-09

Wang, X., Yao, J., Sun, Y. C., and Wood, T. K. (2021). Type VII toxin/antitoxin classification system for antitoxins that enzymatically neutralize toxins. *Trends Microbiol.* 29, 388–393. doi: 10.1016/j.tim.2020.12.001

Wang, P., Yu, Z., Li, B., Cai, X., Zeng, Z., Chen, X., et al. (2015). Development of an efficient conjugation-based genetic manipulation system for *Pseudoalteromonas*. *Microb. Cell Factories* 14:11. doi: 10.1186/s12934-015-0194-8

Wang, P., Zeng, Z., Wang, W., Wen, Z., Li, J., and Wang, X. (2017). Dissemination and loss of a biofilm-related genomic island in marine *Pseudoalteromonas* mediated by integrative and conjugative elements. *Environ. Microbiol.* 19, 4620–4637. doi: 10.1111/1462-2920.13925

Wozniak, R. A., 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

Yao, J., Guo, Y., Wang, P., Zeng, Z., Li, B., Tang, K., et al. (2018). Type II toxin/antitoxin system $ParE_{SO}$/$CopA_{SO}$ stabilizes prophage CP4So in *Shewanella oneidensis*. *Environ. Microbiol.* 20, 1224–1239. doi: 10.1111/1462-2920.14068

Yao, J., Guo, Y., Zeng, Z., Liu, X., Shi, F., and Wang, X. (2015). Identification and characterization of a HEPN-MNT family type II toxin-antitoxin in

Shewanella oneidensis. *Microb. Biotechnol.* 8, 961–973. doi: 10.1111/1751-7915.12294

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fmicb.2022.856884

Phylogenetic Structure and Comparative Genomics of Multi-National Invasive *Haemophilus influenzae* Serotype a Isolates

*Nadav Topaz[1], Raymond Tsang[2], Ala-Eddine Deghmane[3], Heike Claus[4], Thiên-Trí Lâm[4], David Litt[5], Maria Paula Bajanca-Lavado[6], María Pérez-Vázquez[7], Didrik Vestrheim[8], Maria Giufrè[9], Arie Van Der Ende[10], Olivier Gaillot[11,12], Alicja Kuch[13], Martha McElligott[14], Muhamed-Kheir Taha[3] and Xin Wang[15]**

[1] Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States, [2] Vaccine Preventable Bacterial Diseases, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada, [3] Centre National de Référence des Méningocoques, Institut Pasteur, Paris, France, [4] Institute for Hygiene and Microbiology, University of Würzburg, Würzburg, Germany, [5] Respiratory and Vaccine Preventable Bacterial Reference Unit, Public Health England, London, United Kingdom, [6] Haemophilus Influenzae Reference Laboratory, Department of Infectious Disease, National Institute of Health, Lisbon, Portugal, [7] Laboratorio de Referencia e Investigación en Resistencia a Antibióticos e Infecciones Relacionadas con la Asistencia Sanitaria, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain, [8] Norwegian Institute of Public Health, Division of Infection Control and Environmental Health, Oslo, Norway, [9] Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy, [10] Department of Medical Microbiology and Infection Prevention and the Netherlands Reference Laboratory for Bacterial Meningitis, University of Amsterdam, Amsterdam, Netherlands, [11] Service de Bactériologie-Hygiène, CHU Lille, Lille, France, [12] CNRS, INSERM, U1019-UMR 8204, Center for Infection and Immunity, CHU Lille, Lille, France, [13] Department of Epidemiology and Clinical Microbiology, National Medicines Institute, Warsaw, Poland, [14] Irish Meningitis and Sepsis Reference Laboratory, Children's Health Ireland at Temple Street, Dublin, Ireland, [15] Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States*

OPEN ACCESS

Edited by:
Daniel Yero, Universidad Autónoma de Barcelona, Spain

Reviewed by:
Stephen Glen Tristram, University of Tasmania, Australia
Urszula Kosikowska, Medical University of Lublin, Poland

***Correspondence:**
Xin Wang
gqe8@cdc.gov

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

Received: *17 January 2022*
Accepted: *24 February 2022*
Published: *24 March 2022*

Citation:
Topaz N, Tsang R, Deghmane A-E, Claus H, Lâm T-T, Litt D, Bajanca-Lavado MP, Pérez-Vázquez M, Vestrheim D, Giufrè M, Van Der Ende A, Gaillot O, Kuch A, McElligott M, Taha M-K and Wang X (2022) Phylogenetic Structure and Comparative Genomics of Multi-National Invasive Haemophilus influenzae Serotype a Isolates. Front. Microbiol. 13:856884. doi: 10.3389/fmicb.2022.856884

Recent reports have indicated a rise of invasive disease caused by *Haemophilus influenzae* serotype a (Hia) in North America and some European countries. The whole-genome sequences for a total of 410 invasive Hia isolates were obtained from 12 countries spanning the years of 1998 to 2019 and underwent phylogenetic and comparative genomic analysis in order to characterize the major strains causing disease and the genetic variation present among factors contributing to virulence and antimicrobial resistance. Among 410 isolate sequences received, 408 passed our quality control and underwent genomic analysis. Phylogenetic analysis revealed that the Hia isolates formed four genetically distinct clades: clade 1 (n = 336), clade 2 (n = 13), clade 3 (n = 3) and clade 4 (n = 56). A low diversity subclade 1.1 was found in clade 1 and contained almost exclusively North American isolates. The predominant sequence types in the Hia collection were ST-56 (n = 125), ST-23 (n = 98) and ST-576 (n = 51), which belonged to clade 1, and ST-62 (n = 54), which belonged to clade 4. Clades 1 and 4 contained predominantly North American isolates, and clades 2 and 3 predominantly contained European isolates. Evidence of the presence of capsule duplication was detected in clade 1 and 2 isolates. Seven of the virulence genes involved in endotoxin

biosynthesis were absent from all Hia isolates. In general, the presence of known factors contributing to β-lactam antibiotic resistance was low among Hia isolates. Further tests for virulence and antibiotic susceptibility would be required to determine the impact of these variations among the isolates.

Keywords: serotype a, genomics, phylogenetic analysis, invasive isolates, *Haemophilus influenzae*

INTRODUCTION

Haemophilus influenzae (*H. influenzae*) is a gram negative, commensal organism found in the human respiratory tract that can occasionally cause serious disease such as meningitis, pneumonia and bacteremia (Turk, 1984). There are six encapsulated serotypes (Hia, Hib, Hic, Hid, Hie and Hif), as well as unencapsulated *H. influenzae*, referred to as non-typeable *Haemophilus influenzae* (NTHi). Isolates of each encapsulated serotype produce a unique polysaccharide capsule encoded by the *H. influenzae* capsule locus (Potts et al., 2019), whereas NTHi do not possess this gene locus.

Historically, Hib was the primary cause of meningitis in children under 5 years of age, as well as a major cause of other serious infections such as pneumonia and sepsis, but invasive Hib disease has been dramatically reduced in countries that have incorporated Hib conjugate vaccines into immunization schedules (Peltola, 2000). In the post Hib vaccine era, NTHi and non-Hib serotypes have continued to cause invasive disease, as indicated by reports from both the United States (Soeters et al., 2018) and Europe (van Wessel et al., 2011; Whittaker et al., 2017). While Hif has remained the predominant encapsulated serotype causing invasive disease in the United States (Soeters et al., 2018), there have been reports indicating an increase of invasive Hia disease, particularly in North America (Bender et al., 2010; Soeters et al., 2018) and in indigenous populations (Kelly et al., 2011; Bruce et al., 2013; Tsang et al., 2014). Additionally, Hia infection and disease have been reported in other parts of the world, including in some European countries (Giufrè et al., 2017; Deghmane et al., 2019; Heliodoro et al., 2020). As a result, efforts have been undertaken to develop a vaccine targeting Hia (Cox et al., 2017).

The Hia capsule is known to be a major virulence factor, and its structure has been previously well characterized (Ulanova and Tsang, 2014; Potts et al., 2019). The capsule locus contains three regions, regions I and III are shared across the six encapsulated serotypes, and region II contains genes unique to the serotype. Some strains of Hia harbor two copies of the capsule locus, with one copy containing a partial deletion in the *bexA* gene (Kroll et al., 1994). This mutation has been reported to increase production of the capsule polysaccharide and increase virulence (Kroll et al., 1993). Another major virulence factor is the IgA protease; this protein plays a key role in disrupting the defense of the mucosa conferred by human IgA1 and enabling *H. influenzae* to colonize the respiratory tract (Kilian et al., 1996). Two types of IgA proteases have been previously reported, with each type being associated with certain serotypes: type 1 are produced by serotypes a, b, d, and f, while type 2 are produced by serotypes c and e (Mulks et al., 1982). Other key virulence factors in *H. influenzae* include genes involved in adhesion (Guerina et al., 1982; van Ham et al., 1994), iron uptake, immune evasion, and endotoxins (High et al., 2015).

Treatment for *H. influenzae* meningitis typically consists of a third generation cephalosporin such as cefotaxime or ceftriaxone (Kimberlin et al., 2018), or ampicillin after antibiotic susceptibility testing of the isolate. Certain genetic markers in *H. influenzae* have been previously reported to confer resistance to β-lactam antibiotics, including the expression of the bla$_{TEM}$-1 or bla$_{ROB}$-1 β-lactamases (Doern et al., 1997; Hasegawa et al., 2003) and amino acid variations in the penicillin-binding protein 3 (PBP3, encoded by *ftsI*) (Ubukata et al., 2001; Deghmane et al., 2019) which reduce the affinity of this protein to bind to β–lactams such as cephalosporins.

The increase of Hia in North America as a cause of invasive *H. influenzae* disease along with its recent detection in other parts of the world supports the need to better understand the global distribution of this pathogen. We therefore sought to perform a genomic analysis of Hia strains collected from multiple countries in order to identify the major strains in circulation and the factors contributing to their virulence and antimicrobial resistance. Here, we describe the phylogenetic structure of Hia, including placing Hia in the phylogenetic context of other encapsulated serotypes and identifying which Hia strains are circulating in various regions. Additionally, we performed comparative genomics across the Hia strains to identify genetic differences in virulence genes and genes involved in antimicrobial resistance. Finally, we assessed associations between genotype of the Hia strain and clinical disease phenotype and age group.

MATERIALS AND METHODS

Isolate Collection and Whole-Genome Sequencing

The whole genome sequences for 410 invasive Hia isolates were collected from 12 countries, including the United States, Canada, France, Germany, England, Portugal, Spain, Netherlands, Norway, Italy, Poland and Ireland. The isolate counts, time-period and metadata were collected from each participating country and are described in **Table 1**. The isolates sequenced in this study were collected between 1998 and 2019. All isolates were sequenced using Illumina technologies (HiSeq 2500 or MiSeq) with the exception of four isolates from the United States which were sequenced using PacBio as previously described (Retchless et al., 2016) and one isolate from Italy sequenced using Ion Torrent technology (ThermoFisher Scientific). All sequence reads were transferred to the Bacterial Meningitis

TABLE 1 | Sequences received and metadata used in the study.

Country	Sequenced received	Time period	Clinical presentation	Age group
United States	284	1998–2019	Yes	Yes
Canada	30	2006–2011		Yes
France	24	2010–2019	Yes	Yes
Germany	20	2008–2019	Yes	Yes
England	18	2008–2018	Yes	Yes
Portugal	11	1999–2019	Yes	Yes
Spain	8	2014–2019	Yes	
Netherlands	5	2014–2018	Yes	Yes
Norway	3	2017–2019	Yes	Yes
Italy	3	2015–2018	Yes	Yes
Poland	2	2017	Yes	Yes
Ireland	2	2018–2019	Yes	Yes

Lab at the Centers for Disease Control and Prevention for assembly and genomic analysis. The non-PacBio sequence reads for each sample were trimmed using Cutadapt (Martin, 2011) and subsequently mapped against the human reference genome hg38 as well as a collection of common contaminants and any matching sequence reads were removed. The final sequence reads were used to generate a *de novo* assembly using SPAdes 3.7.0 (Bankevich et al., 2012). Following assembly, contigs with a depth of coverage of less than one-tenth the genome-wide coverage were marked as spurious and removed. Each assembly underwent quality control by checking for the expected genome assembly size of 1.8–2.0 mb along with a minimum of 20× average coverage across the genome to ensure confidence for the subsequent genomic analyses. The sequence data for the Hia isolates used in the project are provided under NCBI Bioproject PRJNA766550.

Genomic Analysis

Each genome assembly was annotated using BLAST (Altschul et al., 1997) against a custom database consisting of the PubMLST *Haemophilus influenzae* allele collection (Jolley and Maiden, 2010) supplemented with additional loci for manually curated genes involved in antimicrobial resistance. The ORFs for gene sequences in each assembly were confirmed with by two gene prediction software, Prodigal (Hyatt et al., 2010) and GeneMarkS2 (Lomsadze et al., 2018). The MLST schemes from PubMLST (Jolley and Maiden, 2010) were used to identify the sequence type (ST) of each *H. influenzae* isolate. Additional sequences for invasive non-Hia isolates collected between 1998 and 2019 from the United States were included for the phylogenetic analysis and comprised of the following: 109 Hif, 78 Hie, 44 Hib, 6 Hic, and 5 Hid. The phylogenetic analyses comprising Hia genomes and non-Hia genomes were performed using kSNP v3 (Gardner et al., 2015) and annotated using the Interactive Tree of Life platform (Letunic and Bork, 2016). The protein specific phylogenetic trees were generated by aligning protein sequences using clustal omega (Sievers et al., 2011) and using the subsequent alignment to create a maximum-likelihood phylogenetic tree using RAxML (Stamatakis, 2014). The Single Nucleotide Polymorphism (SNP) difference calculations were performed using custom python scripts with the SNP alignment generated by kSNP as input. Genes involved in virulence were identified using the Virulence Factor Database platform (Liu et al., 2018). For the comparative genomics analyses, custom python scripts were used to create spreadsheets looking at gene presence or absence across clades, identify alleles of genes harboring internal stop codons within specific clades, and to characterize the variant present in antimicrobial resistance genes. The genome-wide association study was performed using the PySeer (Lees et al., 2018) package on 319 invasive Hia isolates for which clinical presentation data was provided. The presence or absence of the capsule duplication was detected by mapping the sequence reads for isolates sequenced using Illumina technologies against their overall genome assembly and only against the capsule region and calculating the difference in average depth of coverage.

Statistical Analysis

The hypergeometric enrichment test was used to calculate significant over-enrichment of major sequence types observed in the study within age groups; this was limited to sequence types that contained at least 30 isolates to ensure sufficient sample size for statistical confidence.

RESULTS

The Hia Isolates Formed Four Major Clades

Among the 410 Hia isolate genome sequences obtained for the study, 76.59% (314/410) were obtained from North American countries, while 23.41% (96/410) were obtained from European countries. The two isolate sequences from Ireland failed our quality control checks and were therefore excluded from further genomic analysis. Phylogenetic analysis revealed that the remaining 408 Hia isolates formed four major clades (**Figure 1A**) consisting of the following: clade 1 (n = 336), clade 2 (n = 13), clade 3 (n = 3) and clade 4 (n = 56). Each of the four clades contained isolates from both North America and Europe; specifically, clade 1 contained similar percentages of isolates from North America and Europe (80.57 and 88.30%, respectively), clades 2 and 3 contained higher percentages of European isolates (8.51 to 1.59%, and 2.13 to 0.32%, respectively), and clade 4 contained a higher percentage of North American isolates (17.52 to 1.06%) (**Supplementary Material 1**). These clades were genetically diverse from one another, with clades 1 and 2 differing by an average of 9,800 SNPs, clades 3 and 4 differing by an average of 7,500 SNPs, and clades 1 and 2 differing from clades 3 and 4 by an average of 16,000 SNPs. Each of these major clades contained one or more predominant sequence types, with ST-56 (n = 125), ST-23 (n = 98) and ST-576 (n = 51), comprising 81.50% (274/336) of clade 1; ST-4 (n = 12) comprising 92.31% (12/13) of clade 2, ST-60 (n = 2) comprising 67% (2/3) of clade 3, and ST-62 (n = 54) comprising 96.40% (54/56) of clade 4. The Hia isolate genomes were compared against other encapsulated serotypes, including Hib, Hic, Hid,

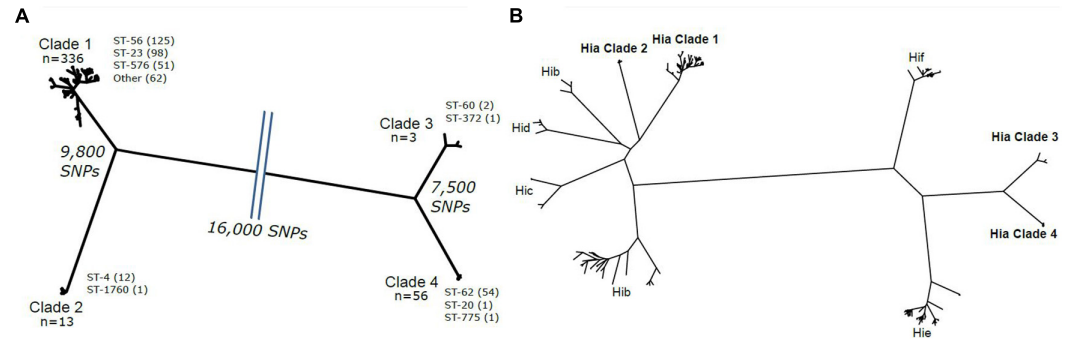


FIGURE 1 | (A) Phylogeny showing the phylogenetic structure of the Hia isolates. Each clade is marked with the number of isolates and the major sequence types and their counts. The SNP difference between each clade is shown. **(B)** The phylogeny is shown including all Hia isolates and the additional encapsulated serotypes. The four Hia clades are shown in bold.

Hif and Hie (**Figure 1B**). A clear delineation was observed with Hia clades 1 and 2 being more closely related to Hib, Hid and Hic, while Hia clades 3 and 4 were more closely related to Hif and Hie.

Most Hia Isolates Belonged to Clade 1

The vast majority (82.3%; 336/408) of the Hia collection belonged to clade 1. The phylogeny for this clade is shown in **Figure 2**. While this clade was diverse with a total of 24 unique STs identified, 16 STs differed by only one MLST allele to one of the three major STs identified in this clade (ST-56, ST-23 and ST-576), with the remaining 5 differing by two MLST alleles. Among the major STs identified within this clade, ST-56 was predominantly found among North American isolates (90.4%; 113/125), while ST-23 was found across isolates collected from both North America (65.3%; 64/98) and Europe (34%, 34/99). Across each of these sequence types, subclades were formed within the tree containing isolates from both Europe and North America, indicating the circulation of these strains across both regions. A low-diversity subclade was formed within this phylogeny and has been labeled as subclade 1.1. The isolates belonging to subclade 1.1 differ by an average of 115 SNPs, as compared to clade 1 overall which differs by an average of 1,300 SNPs. The vast majority of the isolates in subclade 1.1 were obtained from North America (98.1%; 160/163). ST-576 was entirely found within this subclade and was only found among North American isolates. As shown in **Supplementary Material 2**, the clade harboring the ST-576 isolates shares an ancestral lineage with ST-56 and diverged from a most recent common ancestor with ST-56. The earliest ST-576 isolate in our collection was collected in 2006, indicating that ST-576 diverged from ST-56 prior to 2006, and that this strain has been in circulation in North America since.

Genetic Differences in Major Virulence Factors Were Identified Among the Four Clades

All Hia isolates were confirmed to contain the genes encoding the serotype a capsule locus. There were genetic differences identified within the capsule locus among the clades. Most notably, the capsule locus in clades 1 and 2 were flanked by IS1016 and contained two genes encoding a transposase and integrase downstream of *hcsB* that were not present in clades 3 and 4 (**Figure 3**). The presence of the previously reported capsule duplication in Hib was checked in the Hia isolates. As most isolates were sequenced using short read WGS methodologies, the capsule duplication would not be evident in the resulting genome assemblies. As such, a read mapping approach was undertaken to confirm the presence of the capsule duplication in the isolates (**Supplementary Material 3**). All Hia isolates belonging to clades 3 and 4 had similar average depth of coverage of the capsule region as compared to the rest of the overall genome (0.8–1.2× fold difference), indicating that the capsule duplication was likely not present in these isolates. Conversely, the vast majority of isolates belonging to clades 1 and 2 (74.6%; 259/347) had 1.5× or greater average depth of coverage in the capsule region as compared to the entire genome (1.5×–7.2× fold difference), indicating that the capsule duplication is likely present in these isolates. In addition to the capsule locus, the Hia collection was scanned for the sequences of 90 known *H. influenzae* virulence factors as defined by the Virulence Factor DataBase (Liu et al., 2018). Information about the virulence genes identified is provided in **Supplementary Material 4**. The following seven virulence factors were absent from all isolates: HAEM0992 (hypothetical protein), *kfiC*, *orfE*, *siaA*, *wbaP*, *rffG* and *lex2B*. The *hifABCDE* genes were only present in clades 3 and 4, with all isolates in these clades harboring the genes. All isolates in clade 4 (n = 56) contained a disrupted *pilB* gene harboring a point mutation at position 682 C- > T creating a premature internal stop codon. Two distinct *hgpC* genes were identified in 165 isolates belonging to clade 1, with 100 of these belonging to subclade 1.1. The *iga gene* was identified in all Hia isolates but was disrupted in 10.2% (42/408) of the collection. Three distinct indel mutations introducing frameshifts were identified comprising 73% (30/42) of the isolates with disrupted *iga gene*s. Each of these indel mutations are represented in **Figure 4**. These mutations consist of the following: 1121delG (black box), 1122insG (red box) and 1161delA (yellow box). The remaining 27% (17/63) of the disrupted *iga gene*s consisted of various point mutations and

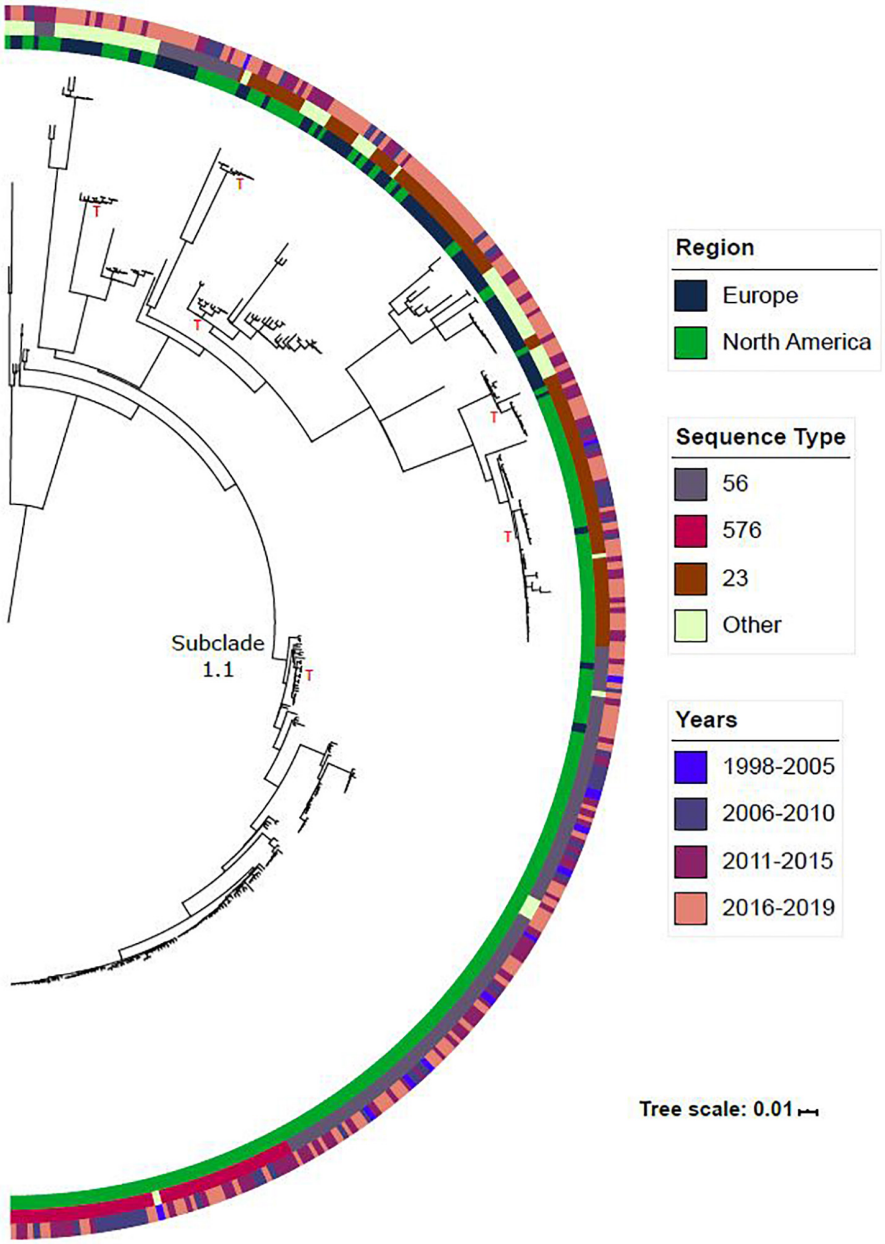


FIGURE 2 | The phylogeny for clade 1 is shown in this figure. Metadata such as region, sequence type and years are included for reference. Other sequence type refers to STs with less than 15 isolates. A red "T" is included next to subclades that contain isolates from both North America and Europe. Subclade 1.1 is labeled in the figure.

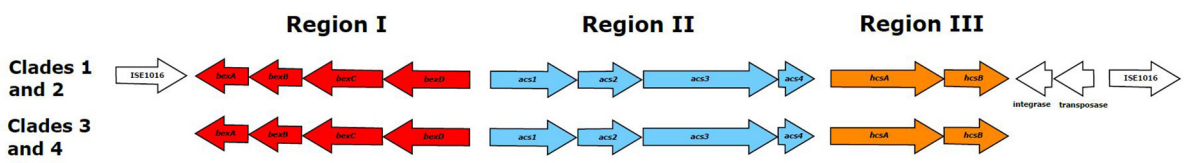


FIGURE 3 | The capsule regions for Hia isolates belonging to each of the four clades is shown. Clades 1 and 2 contain genes encoding an integrase and transposase proteins and are flanked by IS1016 elements.

indels with no distinct pattern. Nearly all isolates belonging to clade 1 and subclade 1.1 (99.4%; 334/336) harbored the outer membrane protein P5 allele 1, while the isolates from other clades harbored different alleles for the P5 protein.

Genetic Elements Associated With Antimicrobial Resistance of Hia

The Hia genome collection was scanned for the presence of genes known to confer resistance to antimicrobials. The *cat* gene was identified in 1.7% (7/408) of the isolates in the collection. All 7 of these isolates belonged to clade 2, ST-4 and were collected from the following countries: United States (n = 3), France (n = 1), England (n = 1), Italy (n = 1) and Netherlands (n = 1). The bla$_{TEM}$*-1* gene was identified in 1.7% (7/408) of the isolate collection. These isolates belonged to the following clades: clade 1 (n = 2), clade 2 (n = 3) and clade 4 (n = 2), and were collected from the following countries: United States (n = 4), England (n = 2) and Canada (n = 1). The bla$_{ROB}$*-1* gene was not identified in any of the Hia isolates.

The diversity of the penicillin-binding protein 3 (PBP3) was assessed across the Hia isolate collection (**Figure 5**). There was a predominant PBP3 variant for each of the major clades. The majority of clade 1 isolates (91%; 306/336) harbored the variant containing amino acid substitutions P31S, L50F, V547I and E603D. None of the clade 2 isolates harbored a variant that was different than the reference. Only one PBP3 variant was detected within clades 3 and 4. Seven European isolates within clade 1 harbored a distinct PBP3 variant containing only amino acid substitution A239E; these isolates consisted of six from England and one from Portugal. Additionally, there were three clade 1 isolates that harbored a PBP3 variant that was more closely related to the main variant in clade 4, and these were collected from the following countries: Norway (n = 1), United States (n = 1) and Canada (n = 1). Finally, three isolates from clade 1, as well as the isolates of clade 4, harbored the D350N substitution that was previously reported in beta-lactamase negative ampicillin resistant (BLNAR) isolates (Deghmane et al., 2019), but has not been reported to confer resistance on its own. However, no other BLNAR-linked substitutions were observed among Hia isolates of this study (**Figure 5**). Furthermore, none of the isolates harbored either of the BLNAR defining N526K or R517H substitutions.

Genetic Association With Clinical and Epidemiological Characteristics

We obtained clinical presentation data for 319 Hia isolates in the collection. These clinical presentations were as follows: 101 pneumonia, 85 meningitis, 73 bacteremia, 24 septicemia, 20 other infections, 6 septicemia and meningitis, 5 arthritis, 4 epiglottitis, and 1 osteomyelitis. A genome wide association study was conducted on these 319 Hia genomes in order to identify regions of the DNA that may be significantly associated with the clinical presentation. Both k-mers and SNPs were considered in this analysis. This analysis revealed a low heritability (h^2) estimate of 0.1, indicating that very little of the phenotypic variation was due to the genetic variation present in this population. Additionally, no k-mers or SNPs were identified to be significantly associated with any of the given clinical phenotypes.

Age group data was available for a total of 336 Hia isolates. These age groups were as follows: < 1, 1–4, 5–10, 11–17, 18–34, 35–49, 50–64 and > 65. Age groups 5–10, 11–17, and 18–34 were merged into one age group (5–34) to provide sufficient sampling for the analysis. A hypergeometric test for enrichment was performed to determine if any of the major sequence types in our study were significantly over-enriched within any of these age groups (**Supplementary Material 5**). The following sequence types with 30 or more isolates that had age group data were included in this analysis: ST-56 (n = 109), ST-23 (n = 77), ST-62 (n = 46) and ST-576 (n = 34) for a total of 266 isolates. No sequence type was identified to be significantly over-enriched within any of the tested age groups.

DISCUSSION

In this study, the genomes for Hia isolates collected from 12 countries were compared in order to better understand the genomic factors of this rising cause of *H. influenzae* disease. We identified four major clades that were distinct from each other and from other encapsulated *H. influenzae* serotypes. Most isolates belonged to clade 1, suggesting that the isolates belonging to clade 1 are the major cause of Hia disease. It would be interesting to further investigate whether the strains belonging to clade 1 are more virulent or transmissible compared to strains of the other Hia clades. The predominant sequence types in clade 1 were ST-56, ST-23 and ST-576. ST-576 was identified in a low diversity subclade within clade 1 and only contained isolates from North America. ST-576 shares an ancestral lineage with ST-56 and diverged from a most recent common ancestor of ST-56 sometime prior to 2006. Obtaining more Hia isolates from other countries would be needed to confirm if this subclade 1.1 is exclusively circulating within North America or in other countries as well.

Previously, two distinct Hia clades have been reported (Potts et al., 2019) which correspond to clades 1 and 4 in our study. The genome sequences belonging to clades 3 and 4 became available more recently, providing additional resolution on the phylogenetic structure of this serotype. While the presence of four major clades suggests that Hia may be more diverse than other encapsulated serotypes, including additional encapsulated *H. influenzae* isolates in our comparison would be needed to confirm this result. Hib contained two distinct clades and was the only other serotype to contain more than one phylogenetically distinct group.

Our collection was largely dominated by isolates collected from North America. This could be explained by the fact that Hia is much less prevalent in Europe as compared to North America (Whittaker et al., 2017), suggesting that our sampling is likely representative of the Hia disease burden in these countries for recent years. Furthermore, we did identify isolates from both North America and Europe in each of the four major clades, as well as several instances of subclades containing isolates from

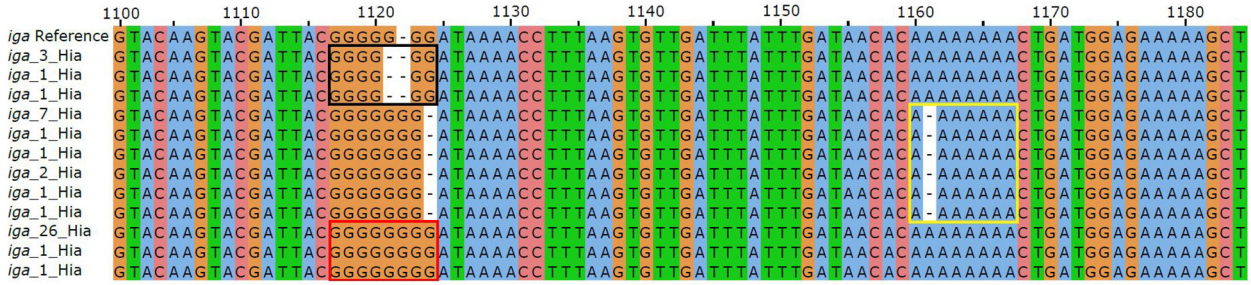


FIGURE 4 | Sequence alignment of the *iga gene* consisting of regions 1,100–1,180 with respect to the reference sequence. The three sites of indels are boxed in black, yellow and red. Each variant is labeled with the count of isolates harboring that variant.

FtsI variants

Clade 1	**Counts**
P31S, L50F, V547I, E603D	306
P31S, L50F, V547I, A595S, E603D	10
A239E	7
P31S, L50F, V547I, A586S, E603D	8
P31S, L50F, G510S, V547I, E603D	1
P31S, L50F, D346G, V547I, E603D	1
S7P, Q28K, V34L, N75S, E135Q, S152A, L165S, S166N, L219I, A239E,D350N, V547I, N569S, A586S, S594T, A595T, E603N	3
Clade 2	
None	13
Clade 3	
S7P, V34L, N75S, E135Q, S152A, L165S, S166N, I180M, L219I, A239E, S273A, V547I, N569S, E603N	3
Clade 4	
S7P, V34L, N75S, E135Q, S152A, L165S, S166N, L219I, A239E, D350N, V547I, N569S, A586S, S594T, A595T, E603N	56

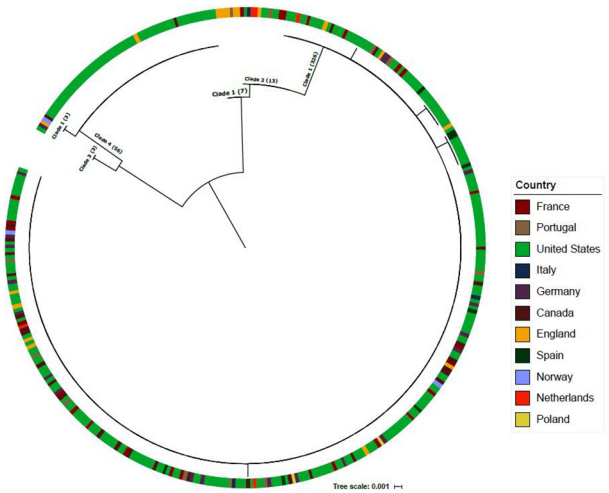


FIGURE 5 | The PBP3 variants are shown in the table for each of the major Hia clades. The phylogeny consists of PBP3 protein sequences and is annotated with the respective country of the isolate. The branches include which of the four major Hia clades the isolates belong to.

both North America and Europe, indicating that these strains have been transmitted across these regions.

It has been observed previously that Hia can cause severe disease similar to Hib (Ulanova and Tsang, 2014; Plumb et al., 2018) in the pre-Hib vaccination era. In our study, we checked for virulence genes that may have been shared among Hia and Hib which could explain this similarity in disease severity but did not identify anything remarkable. We also examined our Hia isolate collection for the presence of the duplicated capsule locus, which has been previously reported to be associated with increased virulence in both Hia and Hib, and found evidence of this mutation being present in the vast majority of the Hia isolates belonging to clades 1 and 2 and absent in the clade 3 and 4 isolates. This finding is supported by the isolates in Hia clades 1 and 2 sharing the same lineage as Hib. Furthermore, we showed that the most predominant Hia strains in circulation, those belonging to clade 1, were most closely related genetically to Hib than the other serotypes. Additionally, it has been previously shown that the Hia and Hib genomic capsular region share stark similarities (Potts et al., 2019), specifically in region II, which encodes biosynthesis genes unique to each serotype. Among these, *acs1* (Hia) and *bcs1* (Hib) are highly similar, sharing a minimum of 96.13% nucleotide sequence similarity (Potts et al., 2019). These factors, including the presence of the capsule duplication and the overall higher level of genetic similarity across the genome and in the capsule region could explain the similar levels of disease severity among Hia and Hib.

Among the four Hia clades, the capsule region of isolates belonging to clades 1 and 2 were flanked by insertion sequence IS1016 and additionally contained two genes encoding a transposase and an integrase. This same pattern of capsule being flanked by IS1016 and containing a transposase and integrase gene was also observed in Hib, Hic and Hid, and seems to be a trait shared among all serotypes in that lineage. Conversely, this pattern was absent from Hia clades 3 and 4, Hif and Hie. This finding is consistent with previous literature describing two phylogenetically distinct divisions of encapsulated *H. influenzae*, one containing capsule regions flanked by IS1016, and the other with IS1016 being absent from the capsule region (Musser et al., 1988; Satola et al., 2008). The presence of a transposase and integrase within the IS1016 flanked

region of the capsule of these genomes suggests that these strains likely acquired the capsule at some point and could be an additional factor in the similar level of disease severity among Hib and Hia.

A total of 90 virulence genes were examined for notable differences between the four major Hia clades. These genes spanned several different classes of virulence factors, such as adherence, endotoxin, immune evasion and iron uptake. Seven of the virulence genes involved in the lipooligosaccharide (LOS) endotoxin biosynthesis were absent from all Hia isolates: HAEM0992 (hypothetical protein), *kfiC*, *orfE*, *siaA*, *wbaP*, *rffG* and *lex2B*. Four of these, *kfiC*, *orfE*, *siaA* and *wbaP*, have been reported previously to be absent in Hia, as well as in other serotypes (Pinto et al., 2019). The pilus genes *pilABCD* were present across all of the four Hia clades, with all isolates in clade 4 harboring a disrupted *pilB* with the same point mutation introducing a premature internal stop. These four genes are required for biogenesis and function of the type IV pilus (Tfp), and each gene has been shown to be required for pilus function (Carruthers et al., 2012). Further experiments would be required to confirm the lack of expression of the PilB protein in clade 4 isolates as a result of this mutation and subsequently assess the potential impact on the virulence of these strains. All Hia isolates across the four clades contained both the *hgpB* and *hgpC* gene, with many isolates in clade 1 harboring two distinct *hgpC* genes. Two other *hgp* genes have been described in the literature, *hgpA* (Morton et al., 1999) and *hgpD* (Harrison et al., 2005), but we did not identify any Hia isolates harboring either of these genes in our collection. The *iga gene* was identified in all Hia isolates, with a small fraction of the isolates (14%) containing genes harboring disrupted ORFs; these disruptions were most commonly introduced by indels in guanine or adenine simple sequence repeat (SSR) regions resulting in frameshift mutations. SSRs serve as an important mechanism for adaption in *H. influenzae* across many genes (Power et al., 2009), and similar mutations in *iga* genes have been previously described in strains of NTHi collected from the airways of patients with Chronic Obstructive Pulmonary Disease (Gallo et al., 2018). Almost all isolates in clade 1 and subclade 1.1 shared an exclusive allele (allele 1) encoding the outer membrane protein P5, while the isolates from other clades harbored different P5 alleles. These different P5 alleles primarily vary by four outer surface loops that are likely involved in the capacity of binding the complement negative regulator, factor H, which is an important inhibitor of alternative pathway activation (Rosadini et al., 2014).

The rise of mutations reported to confer resistance to β-lactam antibiotics is of concern, particularly since this antibiotic class is a primary treatment for *H. influenzae* meningitis (Kimberlin et al., 2018). In general, the presence of known factors contributing to β-lactam resistance were low in our collection, as the bla_{ROB-1} gene was not identified in any of the Hia isolates, and the bla_{TEM-1} gene was only identified in 7 Hia isolates. While we did identify a primary PBP3 variant for each of the four Hia clades, there was genetic variation present in several isolates that was based on region. None of the variants identified have been previously reported to confer resistance, and additional antibiotic susceptibility would be necessary to determine the effects of these specific mutations.

The mechanisms for invasion of *H. influenzae* into the respiratory tract and bloodstream have been previously well described (Clementi and Murphy, 2011; King, 2012) and is dependent on a variety of factors, including those within the host and those within the bacteria. In our study, we did not identify any regions of the Hia genomes to be significantly associated with any of the given clinical disease phenotypes, suggesting that the clinical disease manifestation is a likely a result of the host's response to the pathogen. Among the predominant Hia sequence types in our study, none were significantly more frequent in any age groups considered; however, the highest percentage of the Hia isolates with age data in our study were collected from patients below the age of 4 (150/336; 45%), and the second highest was from patients above the age of 50 (125/336; 37%).

CONCLUSION

In the post-Hib vaccination era, strains of NTHi and encapsulated non-b serotypes have continued to be major causes of *H. influenzae* disease. Among these, Hia is of particular interest given increases in cases reported in North America and in indigenous populations. The results of our study have shown that the reported cases of Hia in North America and Europe in the past two decades have belonged to four genetically distinct clades, with ST-56, ST-23, ST-576, and ST-4 being the predominant sequence types. Variations in virulence genes, such as mutations disrupting the ORFs, as well as the presence/absence and variation of antimicrobial resistance markers were identified within the clades. Additional tests of virulence and antibiotic susceptibility would be needed to determine the impact of these variations among the isolates.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

NT, M-KT, and XW contributed to the study design, data collection, analysis, and critical review and approval of the manuscript. NT prepared the first manuscript draft. All authors have contributed to data collection, investigation, and critical review and approval of the manuscript.

FUNDING

This work was funded by Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

We would like to acknowledge all the country partners who submitted the *H. influenzae* isolate sequencing and metadata used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.856884/full#supplementary-material

REFERENCES

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

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

Bender, J. M., Cox, C. M., Mottice, S., She, R. C., Korgenski, K., Daly, J. A., et al. (2010). Invasive *Haemophilus influenzae* disease in Utah children: an 11-year population-based study in the era of conjugate vaccine. *Clin. Infect. Dis.* 50, e41–e46. doi: 10.1086/651165

Bruce, M. G., Zulz, T., DeByle, C., Singleton, R., Hurlburt, D., Bruden, D., et al. (2013). *Haemophilus influenzae* serotype a invasive disease, Alaska, USA, 1983–2011. *Emerg. Infect. Dis.* 19:932. doi: 10.3201/eid1906.121805

Carruthers, M. D., Tracy, E. N., Dickson, A. C., Ganser, K. B., Munson, R. S. Jr., and Bakaletz, L. O. (2012). Biological roles of nontypeable *Haemophilus influenzae* type IV pilus proteins encoded by the pil and com operons. *J. Bacteriol.* 194, 1927–1933. doi: 10.1128/JB.06540-11

Clementi, C., and Murphy, T. (2011). Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. *Front. Cell. Infect. Microbiol.* 1:1. doi: 10.3389/fcimb.2011.00001

Cox, A. D., Barreto, L., Ulanova, M., Bruce, M. G., and Tsang, R. (2017). Conference c. Developing a vaccine for *Haemophilus influenzae* serotype a: proceedings of a workshop. *Can. Commun. Dis. Rep.* 43, 89–95. doi: 10.14745/ccdr.v43i05a02

Deghmane, A.-E., Hong, E., Chehboub, S., Terrade, A., Falguières, M., Sort, M., et al. (2019). High diversity of invasive *Haemophilus influenzae* isolates in France and the emergence of resistance to third generation cephalosporins by alteration of ftsI gene. *J Infect.* 79, 7–14. doi: 10.1016/j.jinf.2019.05.007

Doern, G. V., Brueggemann, A. B., Pierce, G., Holley, H. P. Jr., and Rauch, A. (1997). Antibiotic resistance among clinical isolates of *Haemophilus influenzae* in the United States in 1994 and 1995 and detection of beta-lactamase-positive strains resistant to amoxicillin-clavulanate: results of a national multicenter surveillance study. *Antimicrob. Agents Chemother.* 41, 292–297. doi: 10.1128/AAC.41.2.292

Gallo, M. C., Kirkham, C., Eng, S., Bebawee, R. S., Kong, Y., Pettigrew, M. M., et al. (2018). Changes in IgA protease expression are conferred by changes in genomes during persistent infection by nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Infect. Immun.* 86:e00313-18. doi: 10.1128/IAI.00313-18

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3. 0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878. doi: 10.1093/bioinformatics/btv271

Giufrè, M., Cardines, R., Brigante, G., Orecchioni, F., and Cerquetti, M. (2017). Emergence of invasive *Haemophilus influenzae* type a disease in Italy. *Clin. Infect. Dis.* 64, 1626–1628.

Guerina, N., Langermann, S., Clegg, H., Kessler, T., Goldmann, D., and Gilsdorf, J. R. (1982). Adherence of piliated *Haemophilus influenzae* type b to human oropharyngeal cells. *J. Infect. Dis.* 146:564. doi: 10.1093/infdis/146.4.564

Harrison, A., Dyer, D. W., Gillaspy, A., Ray, W. C., Mungur, R., Carson, M. B., et al. (2005). Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. *J. Bacteriol.* 187, 4627–4636. doi: 10.1128/JB.187.13.4627-4636.2005

Hasegawa, K., Yamamoto, K., Chiba, N., Kobayashi, R., Nagai, K., Jacobs, M. R., et al. (2003). Diversity of ampicillin-resistance genes in *Haemophilus influenzae* in Japan and the United States. *Microb. Drug Resist.* 9, 39–46.

Heliodoro, C. I. M., Bettencourt, C. R., and Bajanca-Lavado, M. P. (2020). Molecular epidemiology of invasive *Haemophilus influenzae* disease in Portugal: an update of the post-vaccine period, 2011-2018. *Eur. J. Clin. Microbiol. Infect. Dis.* 39, 1471–1480.

High, N. J., Fan, F., and Schwartzman, J. D. (2015). "Chapter 97 – *Haemophilus influenzae*," in *Molecular Medical Microbiology*, 2nd Edn, eds Y.-W. Tang, M. Sussman, D. Liu, I. Poxton, and J. Schwartzman (Boston, MA: Academic Press), 1709–1728.

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

Jolley, K. A., and Maiden, M. C. (2010). BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595

Kelly, L., Tsang, R. S., Morgan, A., Jamieson, F. B., and Ulanova, M. (2011). Invasive disease caused by *Haemophilus influenzae* type a in Northern Ontario First Nations communities. *J. Med. Microbiol.* 60, 384–390. doi: 10.1099/jmm.0.026914-0

Kilian, M., Reinholdt, J., Lomholt, H., Poulsen, K., and Frandsen, E. V. (1996). Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence. *APMIS* 104, 321–338. doi: 10.1111/j.1699-0463.1996.tb00724.x

Kimberlin, D. W. B. M., Jackson, M. A., and Long, S. S. (2018). "*Haemophilus influenzae* infections," in *Red Book: 2018 Report of the Committee on Infectious Diseases American Academy of Pediatrics*, eds D. W. Kimberlin, M. T. Brady, M. A. Jackson, and S. S. Long (Elk Grove Village, IL: American Academy of Pediatrics).

King, P. (2012). *Haemophilus influenzae* and the lung (Haemophilus and the lung). *Clin. Transl. Med.* 1:10.

Kroll, J. S., Moxon, E. R., and Loynds, B. M. (1993). An ancestral mutation enhancing the fitness and increasing the virulence of *Haemophilus influenzae* type b. *J. Infect. Dis.* 168, 172–176. doi: 10.1093/infdis/168.1.172

Kroll, J. S., Moxon, E. R., and Loynds, B. M. (1994). Natural genetic transfer of a putative virulence-enhancing mutation to *Haemophilus influenzae* type a. *J. Infect. Dis.* 169, 676–679. doi: 10.1093/infdis/169.3.676

Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N., and Corander, J. (2018). pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* 34, 4310–4312. doi: 10.1093/bioinformatics/bty539

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2018). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080

Lomsadze, A., Gemayel, K., Tang, S., and Borodovsky, M. (2018). Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Res.* 28, 1079–1089. doi: 10.1101/gr.230615.117

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.1089/cmb.2017.0096

Morton, D. J., Whitby, P. W., Jin, H., Ren, Z., and Stull, T. L. (1999). Effect of multiple mutations in the hemoglobin- and hemoglobin-haptoglobin-binding proteins, HgpA, HgpB, and HgpC, of *Haemophilus influenzae* type b. *Infect. Immun.* 67, 2729–2739. doi: 10.1128/IAI.67.6.2729-2739.1999

Mulks, M. H., Kornfeld, S. J., Frangione, B., and Plaut, A. G. (1982). Relationship between the specificity of IgA proteases and serotypes in *Haemophilus influenzae*. *J. Infect. Dis.* 146, 266–274. doi: 10.1093/infdis/146.2.266

Musser, J. M., Kroll, J. S., Moxon, E. R., and Selander, R. K. (1988). Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7758–7762. doi: 10.1073/pnas.85.20.7758

Peltola, H. (2000). Worldwide *Haemophilus influenzae* type b disease at the beginning of the 21st century: global analysis of the disease burden 25 years

after the use of the polysaccharide vaccine and a decade after the advent of conjugates. *Clin. Microbiol. Rev.* 13, 302–317. doi: 10.1128/CMR.13.2.302

Pinto, M., González-Díaz, A., Machado, M. P., Duarte, S., Vieira, L., Carriço, J. A., et al. (2019). Insights into the population structure and pan-genome of *Haemophilus influenzae*. *Infect. Genet. Evol.* 67, 126–135. doi: 10.1016/j.meegid.2018.10.025

Plumb, I. D., Lecy, K. D., Singleton, R., Engel, M. C., Hirschfeld, M., Keck, J. W., et al. (2018). Invasive *Haemophilus influenzae* serotype a infection in children: clinical description of an emerging pathogen—Alaska, 2002–2014. *Pediatr. Infect. Dis. J.* 37, 298–303. doi: 10.1097/INF.0000000000001764

Potts, C. C., Topaz, N., Rodriguez-Rivera, L. D., Hu, F., Chang, H. Y., Whaley, M. J., et al. (2019). Genomic characterization of *Haemophilus influenzae*: a focus on the capsule locus. *BMC Genomics* 20:733. doi: 10.1186/s12864-019-6145-8

Power, P. M., Sweetman, W. A., Gallacher, N. J., Woodhall, M. R., Kumar, G. A., Moxon, E. R., et al. (2009). Simple sequence repeats in *Haemophilus influenzae*. *Infect. Genet. Evol.* 9, 216–228.

Retchless, A. C., Hu, F., Ouédraogo, A.-S., Diarra, S., Knipe, K., Sheth, M., et al. (2016). The establishment and diversification of epidemic-associated serogroup W meningococcus in the African Meningitis Belt, 1994 to 2012. *mSphere* 1:e00201-16. doi: 10.1128/mSphere.00201-16

Rosadini, C. V., Ram, S., and Akerley, B. J. (2014). Outer membrane protein P5 is required for resistance of nontypeable *Haemophilus influenzae* to both the classical and alternative complement pathways. *Infect. Immun.* 82, 640–649. doi: 10.1128/IAI.01224-13

Satola, S. W., Napier, B., and Farley, M. M. (2008). Association of IS1016 with the hia adhesin gene and biotypes V and I in invasive nontypeable *Haemophilus influenzae*. *Infect. Immun.* 76, 5221–5227. doi: 10.1128/IAI.00672-08

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Soeters, H. M., Blain, A., Pondo, T., Doman, B., Farley, M. M., Harrison, L. H., et al. (2018). Current epidemiology and trends in invasive *Haemophilus influenzae* disease—United States, 2009–2015. *Clin. Infect. Dis.* 67, 881–889.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tsang, R. S., Bruce, M. G., Lem, M., Barreto, L., and Ulanova, M. (2014). A review of invasive *Haemophilus influenzae* disease in the indigenous populations of North America. *Epidemiol. Infect.* 142, 1344–1354.

Turk, D. J. (1984). The pathogenicity of *Haemophilus influenzae*. *J. Med. Microbiol.* 18, 1–16.

Ubukata, K., Shibasaki, Y., Yamamoto, K., Chiba, N., Hasegawa, K., Takeuchi, Y., et al. (2001). Association of amino acid substitutions in penicillin-binding protein 3 with β-lactam resistance in β-lactamase-negative ampicillin-resistant *Haemophilus influenzae*. *Antimicrob. Agents Chemother.* 45, 1693–1699. doi: 10.1128/AAC.45.6.1693-1699.2001

Ulanova, M., and Tsang, R. S. W. (2014). *Haemophilus influenzae* serotype a as a cause of serious invasive infections. *Lancet Infect. Dis.* 14, 70–82. doi: 10.1016/S1473-3099(13)70170-1

van Ham, S. M., van Alphen, L., Mool, F. R., and van Putten, J. P. (1994). The fimbrial gene cluster of *Haemophilus influenzae* type b. *Mol. Microbiol.* 13, 673–684. doi: 10.1111/j.1365-2958.1994.tb00461.x

van Wessel, K., Rodenburg, G. D., Veenhoven, R. H., Spanjaard, L., van der Ende, A., and Sanders, E. A. (2011). Nontypeable *Haemophilus influenzae* invasive disease in The Netherlands: a retrospective surveillance study 2001-2008. *Clin. Infect. Dis.* 53, e1–e7. doi: 10.1093/cid/cir268

Whittaker, R., Economopoulou, A., Dias, J. G., Bancroft, E., Ramliden, M., and Celentano, L. P. (2017). Epidemiology of invasive *Haemophilus influenzae* disease, Europe, 2007-2014. *Emerg. Infect. Dis.* 23, 396–404.

Author Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 March 2022
doi: 10.3389/fmicb.2022.828990

Genome Sequencing of *Rahnella victoriana* JZ-GX1 Provides New Insights Into Molecular and Genetic Mechanisms of Plant Growth Promotion

*Wei-Liang Kong[1,2], Wei-Yu Wang[1,2], Sheng-Han Zuo[1,2] and Xiao-Qin Wu[1,2]**

[1] Co-Innovation Center for Sustainable Forestry in Southern China, College of Forestry, Nanjing Forestry University, Nanjing, China, [2] Jiangsu Key Laboratory for Prevention and Management of Invasive Species, Nanjing Forestry University, Nanjing, China

Genomic information for bacteria within the genus *Rahnella* remains limited. *Rahnella* sp. JZ-GX1 was previously isolated from the *Pinus massoniana* rhizosphere in China and shows potential as a plant growth-promoting (PGP) bacterium. In the present work, we combined the GridION Nanopore ONT and Illumina sequencing platforms to obtain the complete genome sequence of strain JZ-GX1, and the application effects of the strain in natural field environment was assessed. The whole genome of *Rahnella* sp. JZ-GX1 comprised a single circular chromosome (5,472,828 bp, G + C content of 53.53%) with 4,483 protein-coding sequences, 22 rRNAs, and 77 tRNAs. Based on whole genome phylogenetic and average nucleotide identity (ANI) analysis, the JZ-GX1 strain was reidentified as *R. victoriana*. Genes related to indole-3-acetic acid (IAA), phosphorus solubilization, nitrogen fixation, siderophores, acetoin, 1-aminocyclopropane-1-carboxylate (ACC) deaminase, gamma-aminobutyric acid (GABA) production, spermidine and volatile organic compounds (VOCs) biosynthesis were present in the genome of strain JZ-GX1. In addition, these functions were also confirmed by *in vitro* experiments. Importantly, compared to uninoculated control plants, *Pyrus serotina*, *Malus spectabilis*, *Populus euramericana* (Dode) Guinier cv. "San Martino" (I-72 poplar) and *Pinus elliottii* plants inoculated with strain JZ-GX1 showed increased heights and ground diameters. These findings improve our understanding of *R. victoriana* JZ-GX1 as a potential biofertilizer in agriculture.

Keywords: complete genome sequence, *Rahnella victoriana*, biofertilizer, volatile organic compounds, natural field environment

OPEN ACCESS

Edited by:
Baolei Jia,
Chung-Ang University, South Korea

Reviewed by:
Vijay Kumar,
Institute of Himalayan Bioresource Technology (CSIR), India
Andrea Cavalieri,
University of Copenhagen, Denmark
Mahmoud Yaish,
Sultan Qaboos University, Oman

***Correspondence:**
Xiao-Qin Wu
xqwu@njfu.edu.cn

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

Received: *04 December 2021*
Accepted: *07 March 2022*
Published: *30 March 2022*

Citation:
Kong W-L, Wang W-Y, Zuo S-H and Wu X-Q (2022) Genome Sequencing of Rahnella victoriana JZ-GX1 Provides New Insights Into Molecular and Genetic Mechanisms of Plant Growth Promotion. Front. Microbiol. 13:828990. doi: 10.3389/fmicb.2022.828990

INTRODUCTION

According to the latest statistics of the United Nations, the global population is expected to reach 9.7 billion by 2050 (Goswami and Deka, 2020). To meet the growing demand for food, excessive use of chemical fertilizers and pesticides in agricultural production has damaged soil health and ecology (Ruzzi and Aroca, 2015). At the same time, abiotic stresses, including drought, high salinity, and toxic heavy metals are prevalent worldwide (Rosier et al., 2018; Mondal et al., 2021). It is estimated that abiotic stresses affect approximately 10 hectares of land per minute, with three hectares affected

by soil salinization around the world (Ha-Tran et al., 2021). Similar to high salinity, extremely dry conditions lead to an annual harvest loss of 17% of harvests in tropical, arid and semiarid areas (Chandra et al., 2021). Trees suffer even more from the recurrence of abiotic stress owing to their long lifecycle and the global forest cover is gradually decreasing (Teshome et al., 2020). Therefore, appropriate measures need to be taken to make the whole agricultural and forestry ecosystem develop in the direction of sustainable intensification (Pretty et al., 2018).

With progress in the development of genetic tools and technology, our understanding of the "black box" of microorganisms in soil has become more transparent, and we are at a moment when rhizosphere microbial research is likely to be of great application value (Olenska et al., 2020). A large number of studies have determined the important contributions of specific rhizosphere growth-promoting bacteria to sustainable agricultural production. The plant growth-promoting rhizobacteria (PGPR) reported to date are mainly concentrated in *Bacillus*, *Pseudomonas*, *Enterobacter*, *Burkholderia*, *Klebsiella*, *Azospirillum*, and *Serratia* (Liu et al., 2020a; Alexandre et al., 2021; Kusale et al., 2021; Wang et al., 2021; Yasmin et al., 2021). These rhizosphere colonizing microorganisms employ a variety of mechanisms to alleviate adverse soil conditions, thus promoting plant growth. They mobilize soil nutrients by secreting organic acids, siderophores, fixed nitrogen, and dissolved phosphate (Kong et al., 2020a; Liu et al., 2020b; Mariotti et al., 2021). They can also secrete a series of substances, such as 1-aminocyclopropane-1-carboxylate (ACC deaminase), volatiles, spermidine, gamma-aminobutyric acid (GABA), and phytohormones to induce plant resistance to abiotic stress (Zhou et al., 2016b; Liu et al., 2020c; Nascimento et al., 2020; Choudhury et al., 2021). Up to now, PGPR strains have been studied in detail in agricultural crops, but their application in woody plants has not been well-explored.

Strain JZ-GX1 was screened with a high yield phytase as the index in our laboratory and was first classified as *Rahnella aquatilis* based on the 16S rRNA gene sequence. However, with the advancement of taxonomic research, many members of *Rahnella* have been gradually reported, so it is necessary to identify strain JZ-GX1 at a greater resolution. Previous studies have reported that the JZ-GX1 strain exhibited a significant growth-promoting effect on poplar and corn in greenhouse conditions (Li et al., 2013, 2021), so does it have the characteristics of plant growth-promoting in addition to the degradation of phytate? Moreover, the growth-promoting effect of this strain on woody plants in the field needs to be further studied. Therefore, the whole genome sequencing technique was used in this study to provide new insights into the molecular and genetic mechanism of *Rahnella* sp. JZ-GX1 in promoting plant growth. Furthermore, two ecologically important trees and two economic trees were selected for use in field experiments to further explore the growth-promoting ability of strain JZ-GX1 under natural conditions. The overarching goal was to assess this strain as a new microbial inoculant for sustainable agricultural practices.

MATERIALS AND METHODS

Bacterial Strain and DNA Preparation

Rahnella sp. JZ-GX1 was isolated from the rhizosphere soil of a 28-year-old *Pinus massoniana* in Guangxi, China, on April 8, 2011, and was deposited in the Chinese Center for Type Culture Collection (Accession No. CCTCC M2012439). This strain was routinely incubated in Luria-Bertani (LB) liquid media at 28°C with shaking for 24 h. Genomic DNA was extracted using a modified freeze–thawing method (Chen et al., 2020).

Genome Sequencing, Assembly, and Annotation

Rahnella sp. JZ-GX1 genome was sequenced by the single molecule real-time (SMRT) method at Biomarker Technology Co., Ltd. (Beijing, China). The genome assembly was created using Nanopore long reads, and Illumina paired-end sequences were used for base and indel correction. The filtered subreads was assembled using Canu v1.5 software. Finally, the Pilon software is used to further correct the assembled genome using second generation data, and the final genome with high accuracy was obtained.

The predicted gene sequences were compared with the Clusters of Orthologous Groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot, TrEMBL, non-redundant (Nr), and other functional databases by BLAST, and gene functional annotations were obtained (Tatusov et al., 2003). Based on the results from comparison with the Nr database (Aziz et al., 2008), the application software Blast2GO was used to annotate functions according to the GO database. HMMER software was applied for functional annotation based on the Pfam database (El-Gebali et al., 2019). In addition, the functions of genes were annotated and analyzed using COG and KEGG metabolic pathway enrichment analysis.

Phylogenetic Tree Construction and Average Nucleotide Identity Analysis

A phylogenetic tree was constructed by combining the genome of JZ-GX1 with a set of closely related genomes selected from all public KBase genomes using the Insert Genome Into Species Tree 2.1.10 tool.[1] Relatedness was determined by alignment similarity with a selected subset of COG domains. Next, a phylogenetic tree was reconstructed using FastTree (version 2.1.10). Average Nucleotide Identity (ANI) analysis was performed between strain JZ-GX1 and other *Rahnella* isolates included in the phylogenetic tree using an online ANI calculator.[2]

Nucleotide Sequence Accession Numbers

The chromosome, plasmid 1 and plasmid 2 sequences are available under NCBI BioProject PRJNA720502, with accession numbers CP089919, CP089920, and CP089921, respectively.

[1] http://kbase.us

[2] https://www.ezbiocloud.net/tools/ani

Assessment of Indole-3-Acetic Acid and 1-Aminocyclopropane-1-Carboxylate Deaminase Activity

To induce the production of indole-3-acetic acid (IAA), 250 μL of a culture of strain JZ-GX1 grown overnight in LB medium was transferred into 25 mL of TSB medium supplemented with sterile-filtered L-tryptophan (500 μg/mL). The liquid culture was grown at 28°C and 180 rpm, and cells were separated from the exhausted medium by centrifugation (10,000 × *g* for 15 min). The concentration of IAA in the bacterial solution was sampled and determined at 96 h after inoculation. The collected supernatant was filtered through 0.22 μm cellulose acetate filters (DISMIC®; Frisenette ApS, Knebel, Denmark). The bacterial supernatant (1 mL) was mixed with 4 mL of Salkowski's reagent (50 mL of 35% $HClO_4$, 1 mL of 0.5 M $FeCl_3$) and allowed to rest for 30 min in the dark. After incubation, absorbance was measured at 530 nm (pink color) (T60 UV-VIS Spectrophotometer; PG Instruments, Leicester, United Kingdom) and quantified using the calibration curve of an IAA standard with linear regression analysis (Kang et al., 2020). The ACC deaminase activity was determined by previously described method (Penrose and Glick, 2003). The experiment was repeated two times, and each treatment was conducted in triplicate.

Evaluation of Siderophore Production, Nitrogen Fixation, Phosphorus, and Potassium Solubilizing Ability

For siderophores production, strain was cultivated in a Chrome Azurol-S agar assay (Schwyn and Neilands, 1987), and positive results were indicated by the formation of a clear orange zone around the colonies. *Burkholderia pyrrocinia* JK-SH007 was used as a positive control.

Nitrogen fixation ability using N-free Ashby medium agar plates contained 5 g of glucose, 5 g of mannitol, 0.1 g of $CaCl_2{\cdot}2H_2O$, 0.1 g of $MgSO_4{\cdot}7H_2O$, 0.005 g of $Na_2MoO_4{\cdot}2H_2O$, 0.9 g of K_2HPO_4, 0.1 g of KH_2PO_4, 0.01 g of $FeSO_4{\cdot}7H_2O$, 5.0 g of $CaCO_3$, 15.0 g of agar in 1 L of distilled water, the final pH was adjusted to 7.3 (Liu et al., 2017).

The rhizosphere bacteria were assessed for potential inorganic phosphate solubilization on the National Botanical Research Institute's phosphate (NBRIP) growth medium [per liter: 10.0 g of glucose, 5.0 g of $MgCl_2{\cdot}6H_2O$, 0.25 g of MgSO4·$7H_2O$, 0.2 g of KCl and 0.1 g of $(NH_4)_2SO_4$] supplemented with $Ca_3(PO_4)_2$ at a final concentration of 0.5% (Chandran et al., 2014).

Organophosphorus detection medium contained 0.1 g of Calcium phytate, 10.0 g of Glucose, 5.0 g of $MgCl_2$, 0.2g of KCl, 0.25 g of $MgSO_4$, 0.1 g of $(NH_4)_2SO_4$, 16.0 g of agar in 1 L of distilled water, the final pH was adjusted to 7.3 (Shen et al., 2016).

Potassium dissolving ability using modified Aleksandrov medium including 5.0 g of glucose, 0.5 g of $MgSO_4{\cdot}7H_2O$, 0.1 g of $CaCO_3$, 0.006 g of $FeCl_3$; 2.0 g of $Ca_3(PO_4)_2$, 3.0 g of potassium aluminium silicate and 20.0 g agar in 1 L of distilled water, the final pH was adjusted to 7.2 (Etesami et al., 2017). The JZ-GX1 colonies were stabbed in triplicate on all agar plates using sterile toothpicks and positive results were expressed by the presence of solubilization zone around bacterial colony after 120 h of incubation at 28°C.

Each plate detection was composed of nine replicates, and the experiment was repeated twice.

Voges-Proskauer Test for Detection of Acetoin

The strain was inoculated in Methyl Red-Voges-Proskauer (MR-VP) broth (glucose 0.5 g, K_2HPO_4 0.2 g, water 1,000 mL, pH 7.2-7.4). After shaker culture at 30°C for 2 days, the culture medium was mixed with 40% NaOH, a small amount of creatine was added and the mixture was heated in a boiling water bath to test for a positive reaction (shown as a red color) (Rath et al., 2018). *Escherichia coli* DH5α was used as a negative control. Each treatment was composed of three replicates, and the experiment was repeated twice.

Detection of Key Genes Related to Plant Growth-Promoting Traits in JZ-GX1

Genomic DNA was extracted using a modified freeze–thawing method (Chen et al., 2020). Then, polymerase chain reaction (PCR) assays were used to detect the *acdS*, *gadB*, *gadD*, *gabT*, *speA*, *speD*, *G1pase*, *pstA*, *pstC*, and *nirB* genes by designing PCR primers based on the genome sequence of JZ-GX1 strain. The primers used in the experiment are listed in **Supplementary Table 1**.

Plant-Bacterial Dual Growth Experiments

Arabidopsis thaliana Col-0 was used in this study. Seeds were surface sterilized and sown on 1/2 MS agar medium. The Petri dishes were positioned vertically in a growth chamber under a long photoperiod (16 h light/8 h dark) and 70% relative humidity at 22°C. Seven days later, five seedlings were transferred to a new 1/2 MS agar medium, and at the same time, 10 μL of bacterial solution was added to the other compartment, and LB medium without bacterial solution was used as the control. After the bacterial solution was homogeneously distributed in the culture medium, the Petri dish was sealed with parafilm and grew vertically in the same conditions. After 14 days, *A. thaliana* seedlings were removed and weighed with a 1/10,000 balance; the length of the main root of *A. thaliana* was measured and recorded with a Vernier calliper. Then, the seedlings were placed in a Petri dish filled with clear water so that the roots could be fully elongated, and the number of lateral roots was counted and recorded. The root hair of *A. thaliana* was photographed with a Zeiss stereomicroscope (Zeiss Microscope System Standard 16; Carl Zeiss Ltd.; Wetzlar, Germany) (Bhattacharyya et al., 2015; Perez-Flores et al., 2017; Li F. et al., 2020). Each treatment was composed of ten replicates, and the experiment was repeated twice.

Antagonistic Assay of the Volatile Organic Compounds Produced by *Rahnella* sp. JZ-GX1 Against Plant Pathogenic Fungi

The antifungal effect of the volatile organic compounds (VOCs) produced by *Rahnella* sp. JZ-GX1 was evaluated using I plates (90 mm in diameter). I plates are plastic Petri dishes with a partition in the center that divides the plate into two parts. The microorganisms grown on side thus cannot spread to the other side. However, gas exchange can proceed normally. One compartment of the I plates, containing LB agar medium, was inoculated with *Rahnella* sp. JZ-GX1. The other compartment of the I plates, containing PDA medium, was inoculated with pathogenic fungi using mycelial discs (6 mm diameter). I plates with pathogenic fungi on one half served as the control. The fungal plant pathogens used in this study were preserved in the Forest Pathology Laboratory of Nanjing Forestry University. All plates were sealed with Parafilm and cultured for 5–7 days at 25°C (Zhang et al., 2019c, Zhang et al., 2021). Each treatment was composed of three replicates, and the experiment was repeated twice.

Field Experiment

The experiment was conducted at nurseries in Changzhou (31°47′N, 119°58′E) and Suqian (33°72′N, 118°68′E), Jiangsu Province. The areas have a subtropical monsoon climate, with an average annual precipitation of approximately 1,056 mm and an annual average temperature of approximately 15°C. The soil type is paddy soil. Two-year-old *Pyrus serotina*, *Malus spectabilis*, and I-72 poplar and 1-year-old *Pinus elliottii* were selected as indicator tree species. Three replicates were applied following in a randomized block design. Each plant was irrigated with 800 mL of bacterial liquid (including 50 mL of bacterial suspension), and seedlings that were not inoculated with JZ-GX1 were used as controls, with 90 plants per treatment. The whole experiment was carried out in open air, with drip irrigation provided when the weather was dry. After 8 months, plant height and ground diameter were recorded.

Statistical Analysis

Analysis of variance (ANOVA) followed by Duncan's multiple comparison test was performed for comparisons among all means, and Student's *t*-test was performed for comparison of pairs of means with SPSS 22.0 software (IBM Inc., Armonk, NY, United States). Error bars show the standard deviation and significance was defined as $p < 0.01$. Graphs were generated using GraphPad Prism 8.0 (GraphPad Software, Inc., United States).

RESULTS

General Genome Features and Phylogenetic Analysis of Strain JZ-GX1

The complete genome of *Rahnella* sp. JZ-GX1 contained one gapless circular chromosome of 5,472,828 bp and two plasmids, with a G + C content of 53.53%. Twenty two rRNAs, and 77 tRNAs were found in the chromosome, including 0 pseudogenes, 12 CRISPR sequences, 12 gene islands, and two pre-bacteriophages (**Figure 1**). A total of 4,483 protein-coding genes were predicted, of which 88.85% were annotated to

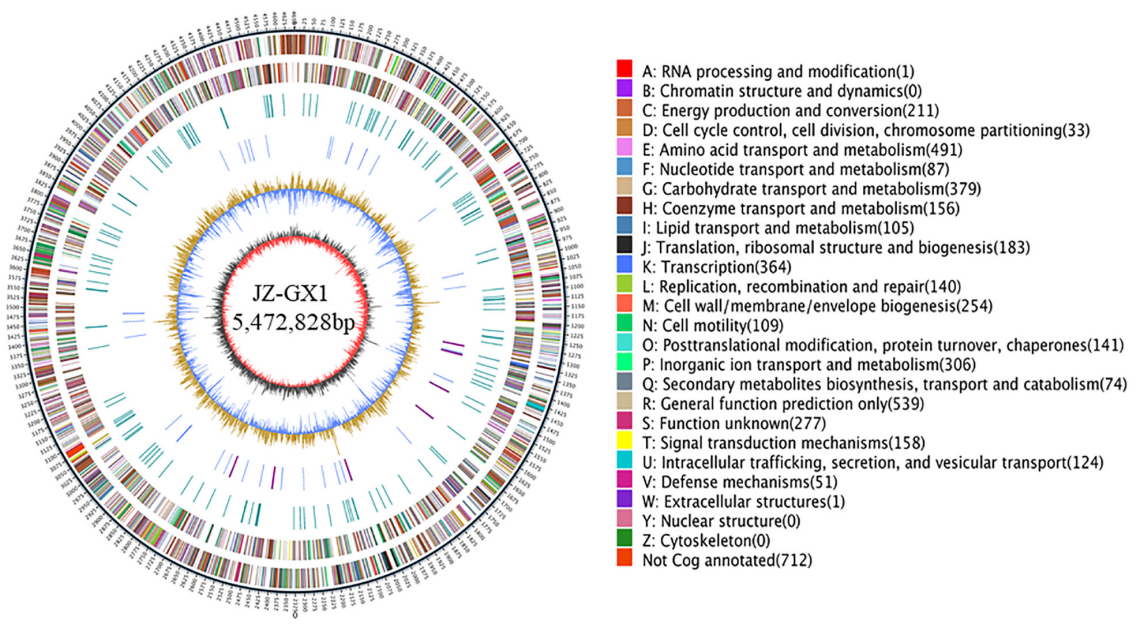


FIGURE 1 | The whole genome of *Rahnella victoriana* JZ-GX1. The genome map is composed of seven circles. From the outer circle to inner circle, each circle displays information regarding the genome of (1) forward CDS, (2) reverse CDS, (3) forward COG function classification, (4) reverse COG function classification, (5) nomenclature and locations of predictive secondary metabolite clusters, (6) G + C content, and (7) GC skew.

COG functional categories, 80.68% to GO functional categories and 68.77% to KEGG pathways. Details are presented in **Supplementary Table 2**.

Phylogenetic analysis based on the whole genome further confirmed the close taxonomic relationship of strain JZ-GX1 with *Rahnella victoriana* (**Figure 2**). On the basis of ANI, the similarity between genomes of strain JZ-GX1 and other *R. victoriana* strains was approximately 98%; when the strain was compared with other species, the ANI threshold was less than 90% (**Table 1**). All the data indicated that strain JZ-GX1 belongs to *R. victoriana* rather than *R. aquatilis*, as formerly proposed.

Gene Function Annotation of *Rahnella victoriana* JZ-GX1

Based on the COG analysis, the identified proteins were classified into 25 functional categories. Among these proteins, in addition to the high content of COG R (general function), the content of COG E (amino acid transport and metabolism) was the highest, followed by COG K (transcription) and COG G (heredity and basic metabolism), indicating that the functional proteins encoded by the cell genome play an important role in maintaining cellular life and genetic metabolism (**Figure 3A**).

Using gene KEGG annotations, 20 functional classes were predicted in the genome of *R. victoriana* JZ-GX1, which can be divided into five categories: (1) systems related to cell processes, such as cell movement (68); (2) systems related to environmental information processing, such as signal transduction (139) and transmembrane transport (340); (3) systems related to genetic information processing; (4) systems related to immunity (39); and (5) systems related to metabolism. Among them, metabolism-related systems accounted for the highest proportion of *R. victoriana* JZ-GX1 metabolic pathways (**Figure 3B**). These results showed that JZ-GX1 strain has strong genetic potential to synthesize secondary metabolites, which is consistent with the above protein annotation results.

Identification of Genes Responsible for Plant Growth-Promoting Property of *Rahnella victoriana* JZ-GX1

Functional analysis of *R. victoriana* JZ-GX1 genome revealed the presence of several genes contributing directly to plant hormones and nutrient availability. We identified 11 genes encoding key enzymes involved in the synthesis and secretion of IAA through the IPyA (*ipdC*) and IAM (*amiE*) pathways. For the N cycle, genes encoding nodulation protein (*nodN*), nitrate reductase catalytic subunit (*nasA*), and nitrite transporter (*nirC*) were detected. These enzymes are involved in the denitrification process, catalyzing the conversion of nitrate to nitrite to nitric oxide, followed by the conversion of nitric oxide to nitrous oxide. Strain JZ-GX1 possessed genes encoding the synthesis of enterobactin and rhizobactin, which form the main part of catechol and hydroxamic acid type siderophores, and a ferric enterobactin transport system (*fepACDG*) was detected. The annotation information also revealed the presence of several gene clusters involved in mineral phosphate solubilization, including the phosphate-specific transport operon (*pstAC*) as well as a two-component signal transduction system for phosphate uptake consisting of *phoPRH* genes (**Table 2**).

Gene Mining for Improving Plant Stress Resistance by *Rahnella victoriana* JZ-GX1

In addition to the genes that directly promote growth, many genes responsible for synthesis and transport of compatible solutes were predicted, such as Na^+/H^+ antiporter (*NhaAB*), glycine betaine transporter (*opuD*), proline/betaine

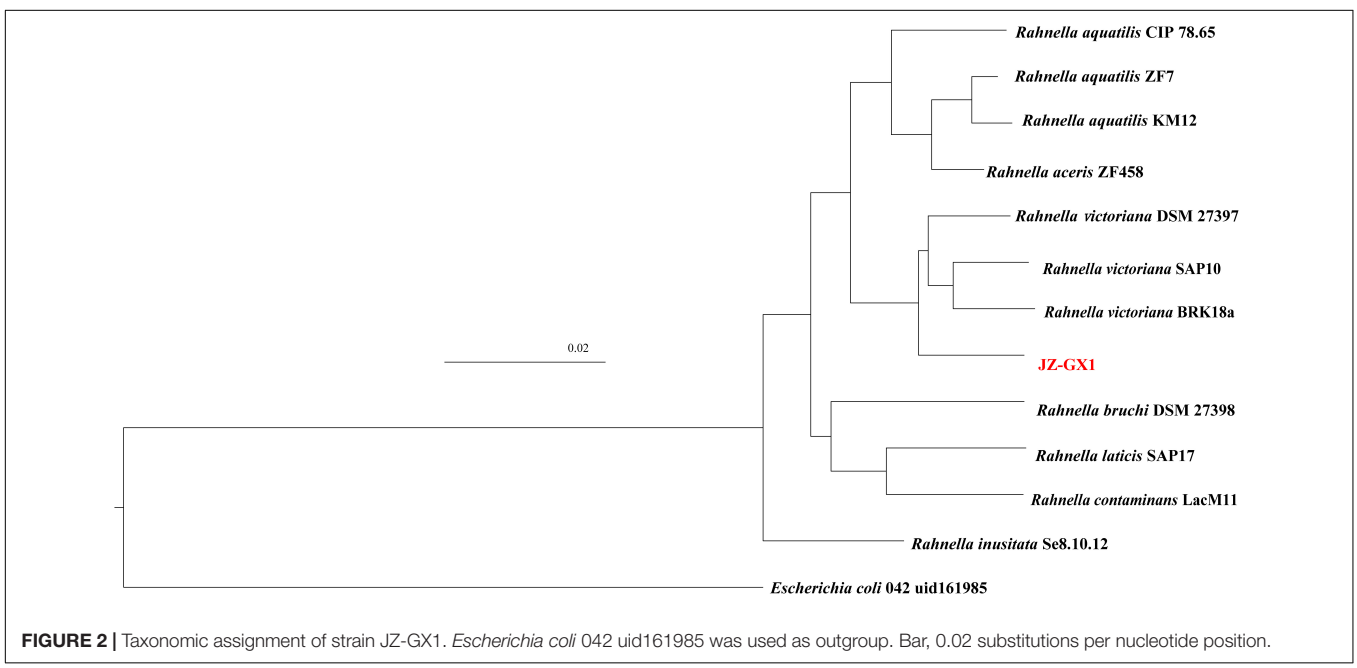


FIGURE 2 | Taxonomic assignment of strain JZ-GX1. *Escherichia coli* 042 uid161985 was used as outgroup. Bar, 0.02 substitutions per nucleotide position.

TABLE 1 | Average nucleotide identity (ANI) values based on alignment of the whole genome of strain JZ-GX1 and the most closely related members of the genus *Rahnella*.

Subject strain	Genome size (bp)	Biosample accession	ANI threshold (%)
JZ-GX1	5,472,828	SAMN18652319	/
R. victoriana SAP-10	5,489,796	SAMN16733768	98.87
R. victoriana DSM 27397	5,309,719	SAMN10987324	98.92
R. victoriana BRK18a	5,563,295	SAMN05249967	98.78
R. aquatilis CIP 78.65	5,448,900	SAMN12024752	89.10
R. aquatilis KM12	4,878,627	SAMN10579152	89.03
R. aquatilis ZF7	5,536,721	SAMN10032142	88.90
R. aceris ZF458	5,602,983	SAMN17207067	88.93
R. bruchi DSM 27398	5,501,702	SAMN10095204	87.07
R. contaminans Lac-M11	5,230,797	SAMN13909976	85.58
R. inusitata Se8.10.12	5,471,205	SAMN16814261	84.45
R. laticis SAP-17	5,727,497	SAMN16734072	85.28

transporter (*proP*), and exopolysaccharide production protein. Genes encoding acetolactate synthase (*alsD* and *alsS*) and 2,3-butanediol (*bdh* and *budABC*), which are involved in the synthesis of acetoin, a volatile molecule associated with plant growth promotion, were detected in the genome. The gene *gadB* responsible for GABA production, as well as *gabT* and *gabD*, encoding GABA aminotransferase and succinate-semialdehyde dehydrogenase involved in GABA degradation, were predicted. Moreover, arginine decarboxylase (*speA*), agmatinase (*speB*), S-adenosylmethionine decarboxylase proenzyme (*speD*) and spermidine synthase (*speE*), related to spermidine biosynthesis, were present in *R. victoriana* JZ-GX1 (**Table 3**).

Plant Growth-Promoting Rhizobacteria Traits Found in *Rahnella victoriana* JZ-GX1 Genome Are Expressed *in vitro*

To verify the accuracy of genome sequencing, we investigated the PGP characteristics of strain JZ-GX1 *in vitro*. The results demonstrated that *R. victoriana* JZ-GX1 appeared as a yellow halo on the CAS plate and Aleksandrov plate, indicating that it can secrete siderophores and dissolve potassium (**Figures 4A,F**). The dissolution circle method showed that JZ-GX1 has the ability to degrade insoluble organic phosphorus and inorganic phosphorus at the same time (**Figures 4B,C**). Moreover, it could grow on Ashby medium, which proved that it has a

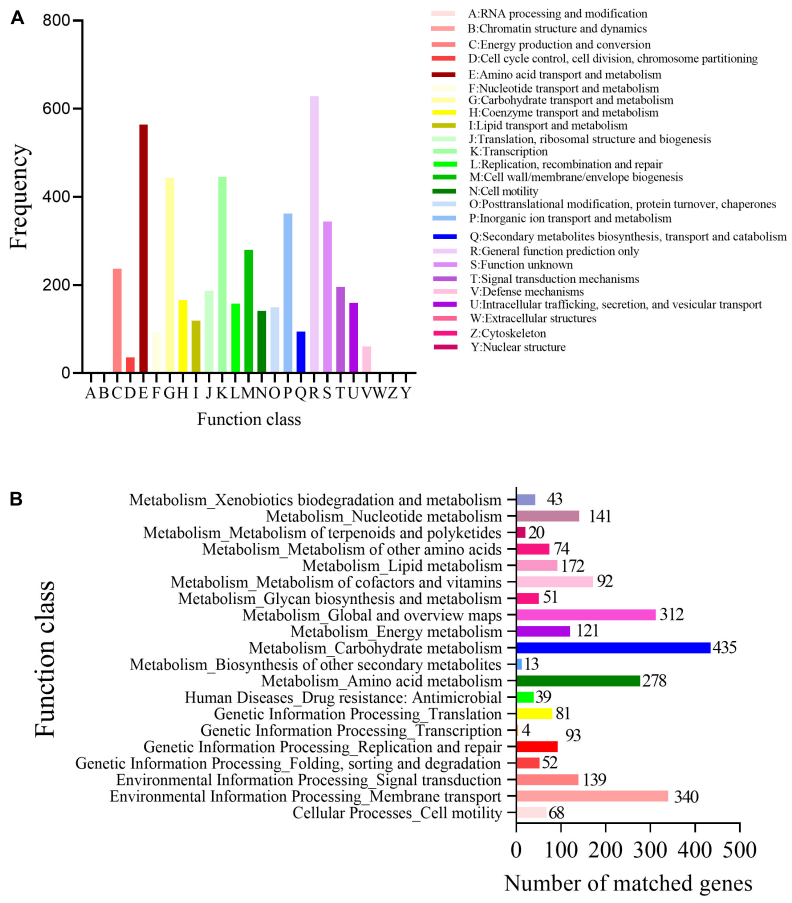


FIGURE 3 | Functional categories of *R. victoriana* JZ-GX1. **(A)** Clusters of Orthologous Groups of proteins (COGs) annotation and **(B)** Clusters of kyoto encyclopedia of genes and genomes (KEGG) annotation.

TABLE 2 | Genes related to plant growth-promoting activities in the *Rahnella victoriana* JZ-GX1 genome.

Gene ID	Gene name	Gene function
Auxin biosynthesis		
GE03900	*trpE*	Anthranilate synthase component I
GE03901	*trpD*	Anthranilate phosphoribosyltransferase
GE03903	*trpB*	Tryptophan synthase subunit beta
GE03904	*trpA*	Tryptophan synthase alpha chain
GE01768	*trpS*	Tryptophanyl-tRNA synthetase
GE00882	*trpR*	Trp operon repressor
GE04238	*mtr*	Tryptophan permease
GE02592	*ipdC*	Indole-3-pyruvate decarboxylase
GE01577	*amiE*	Aliphatic amidase
GE02518	*aec*	Auxin efflux carrier family protein
GE03903	*TSB*	Phosphoribosylanthranilate isomerase
Nitrogen metabolism		
GE00696	*nodN*	Nodulation protein
GE03474		Nitrogen fixation
GE01773	*nirB*	Nitrite reductase (NADH) large subunit
GE01772	*nirD*	Nitrite reductase (NADH) small subunit
GE03448	*nasA*	Nitrate reductase catalytic subunit
GE03466	*nrtP*	MFS transporter, NNP family, nitrate/nitrite transporter
GE04729	*nirC*	Nitrite transporter
Siderophores		
GE02035	*fes*	Enterobactin esterase
GE00216	*entB*	Isochorismatase
GE02737	*entC*	Isochorismate synthase
GE02242	*fepA*	TonB-dependent receptor
GE02036	*fepC*	Ferric enterobactin transport ATP-binding protein
GE02038	*fepD*	Ferric enterobactin transport system permease protein
GE02037	*fepG*	Ferric enterobactin transport system permease protein
GE03776	*RhbF*	Rhizobactin siderophore biosynthesis protein
GE03029	*CirA*	Catecholate siderophore receptor
GE00738	*FhuD*	Fe^{3+}-hydroxamate-binding protein
GE03426	*FhuF*	Ferric iron reductase protein
Phosphorus metabolism		
GE00522	*PhoB*	Phosphate regulon transcriptional regulatory protein
GE03390	*PhoH*	Phosphate starvation-inducible protein
GE00521	*PhoR*	Phosphate regulon sensor protein
GE01248	*PstA*	Phosphate transport system permease protein
GE01484	*PstC*	Phosphate transport system permease protein
GE03803	*pqqA*	Alkaline phosphatase
GE04682	*pqqB*	Pyrroloquinoline quinone biosynthesis protein
GE04681	*pqqC*	Pyrroloquinoline quinone biosynthesis protein
GE04680	*pqqD*	Pyrroloquinoline quinone biosynthesis protein
GE04679	*pqqE*	Pyrroloquinoline quinone biosynthesis protein
GE04678	*pqqF*	Pyrroloquinoline quinone biosynthesis protein
GE01596		4-Phytase activity
GE00546	*G1Pase*	Glucose-1-phosphatase

TABLE 3 | Genes related to plant stress resistance improving in the *Rahnella victoriana* JZ-GX1 genome.

Gene ID	Gene name	Gene function
Salt tolerance		
GE00864	*NhaA*	Na^+/H^+ antiporter
GE03750	*NhaB*	Na^+/H^+ antiporter
GE02187	*opuD*	Glycine betaine transporter
GE02874	*dhaS*	Betaine-aldehyde dehydrogenase
GE02188	*opuBA*	Glycine/betaine ABC transporter ATP-binding protein
GE04729	*proP*	Proline/betaine transporter
GE03756	*eps*	Exopolysaccharide production protein
ACC deaminase		
GE03645	*acds*	1-aminocyclopropane-1-carboxylate deaminase activity
GABA production		
GE04695	*gadB*	glutamate decarboxylase
GE04546	*gabD*	Succinate-semialdehyde dehydrogenase
GE04524	*gabT*	Gamma-aminobutyrate: alpha-ketoglutarate aminotransferase
Spermidine biosynthesis		
GE00638	*SpeA*	arginine decarboxylase
GE00638	*SpeB*	Biosynthetic arginine decarboxylase
GE00775	*SpeD*	S-adenosylmethionine decarboxylase proenzyme
GE00774	*SpeE*	Spermidine synthase
Volatile organic compounds		
GE00932	*alsD*	Alpha-acetolactate decarboxylase
GE00822	*alsS*	Acetolactate synthase
GE04344	*bdh*	(R, R)-butanediol dehydrogenase activity
GE00932	*budA*	Alpha-acetolactate decarboxylase
GE00933	*budB*	Acetolactate synthase, catabolic
GE04344	*budC*	Diacetyl reductase [(S)-acetoin forming]

certain autogenous nitrogen fixation ability (**Figure 4E**). The VP test showed that *R. victoriana* JZ-GX1 could produce acetoin (**Figure 4D**). In addition, the IAA content and ACC deaminase activity were 9.0934 μg/mL and 17.9794 mmol α-keto/mg protein/h, respectively (**Figure 4G**).

We tested the JZ-GX1 strain for the presence of operons for the biosynthesis of ACC deaminase, GABA, and spermidine, the fixation of nitrogen and the dissolution of phosphorus by PCR using specific primers. A fragment of the predicted size for each of these compounds was observed in the DNA of the genome sequence, which indicates that the strain produces these compounds. Fragments of the predicted sizes for *acdS* (∼1,002 bp), *gadB* (∼1,473 bp), *gadD* (∼1,392 bp), *gabT* (∼1,269 bp), *speA* (∼1,983 bp), *speD* (∼795 bp), *G1pase* (∼1,302 bp), *pstA* (∼846 bp), *pstC* (∼957 bp), and *nirB* (∼2,550 bp) were amplified from strain JZ-GX1 DNA (**Figure 4H**).

Volatile Organic Compounds Produced by *Rahnella victoriana* JZ-GX1 Can Promote Plant Growth and Inhibit Phytopathogenic Fungi

To investigate the effect of VOCs produced by *R. victoriana* JZ-GX1 on plant growth, we performed I-plate assays as these avoid any physical contact between *A. thaliana* and *R. victoriana* JZ-GX1. Biomass production was stimulated in *Arabidopsis* plants exposed to *R. victoriana* JZ-GX1 in this manner (**Figure 5A**). Following exposure to *R. victoriana* JZ-GX1, the fresh weights

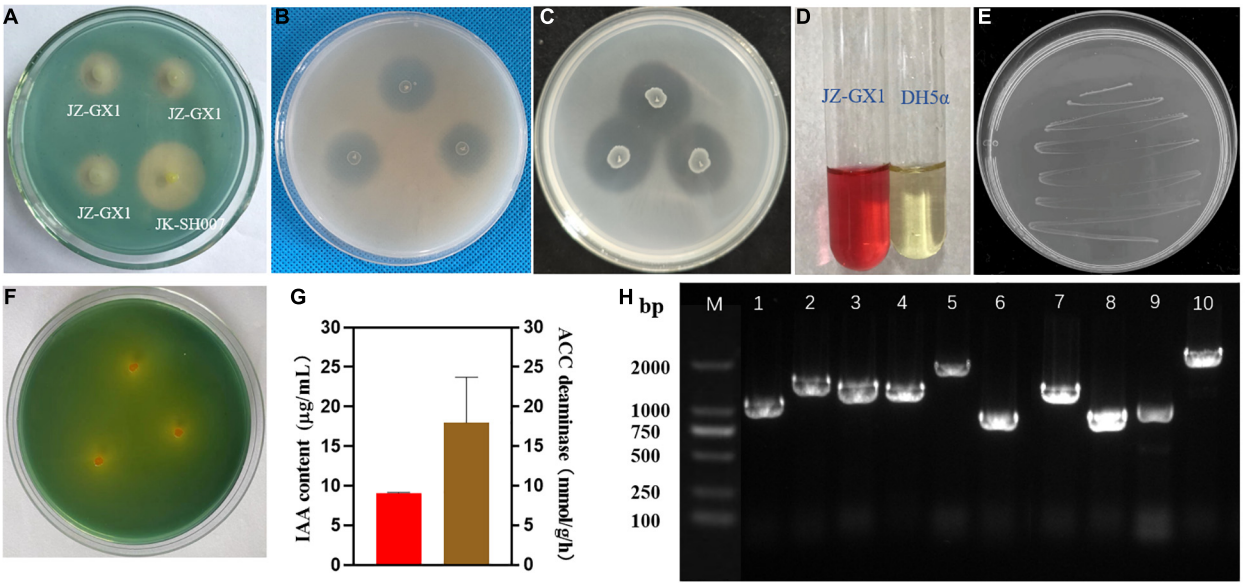


FIGURE 4 | Detection of plant growth-promoting characteristics of *Rahnella victoriana* JZ-GX1. **(A)** Siderophore production, **(B)** inorganic phosphorus dissolution, **(C)** organophosphorus dissolution, **(D)** acetoin production, **(E)** nitrogen fixation, **(F)** potassium solubilization, **(G)** IAA and ACC deaminase production, and **(H)** detection of key gene related to PGP traits in JZ-GX1 by PCR amplification (lane M, DNA marker; 1, *acdS*; 2, *gadB*; 3, *gadD*; 4, *gabT*; 5, *speA*; 6, *speD*; 7, *G1pase*; 8, *pstA*; 9, *pstC*; 10, *nirB*). One-way ANOVA was performed, and Duncan's *post-hoc* test was applied. Different letters indicate statistically significant differences ($p < 0.01$) among treatments ($n = 3$).

of shoots and roots increased significantly by 1.71- and 1.68-fold, respectively, compared to the controls (**Figures 5B,C**). In addition, *Arabidopsis* exposed to strain JZ-GX1 VOCs showed denser root hairs (**Figure 5D**). The bacterial VOCs significantly inhibited primary root growth and stimulated lateral root formation at the same time (**Figures 5E,F**). These data indicate that VOCs from JZ-GX1 can promote plant growth and modulate rhizogenesis in *A. thaliana*.

The VOCs emitted by PGPR often show an inhibitory effect on pathogens. As shown in **Figure 6**, strain JZ-GX1 VOCs showed significant antifungal activities against six plant pathogenic fungi, namely, *Guignardia camelliae*, *Fusicoccus aesculin*, *Fusarium oxysporum*, *Alternaria tenuissima*, *Sphaeropsis sapinea*, and *Verticillium dahlia*. The fungi not exposed to *R. victoriana* JZ-GX1 were covered with plates, while the hyphae of all test fungi except *V. dahlia* in the presence of VOCs were limited to one side of the partition plate, and the pigments of *F. oxysporum* and *V. dahlia* were inhibited.

Growth-Promoting Effect of *Rahnella victoriana* JZ-GX1 on Different Forest Trees in Field Trials

Eight months after inoculation, the microbial agent had a marked growth-promoting effect on the height and ground diameter of I-72 poplar compared with that of the CK, with growth rates of 9.7 and 16.1%, respectively. The growth rates of plant height and ground diameter of *P. elliottii* inoculated with strain JZ-GX1 were 23.6 and 14.6% greater than those of uninoculated trees, respectively. The height of pear trees with root application of strain JZ-GX1 was 14% higher than that of the control trees, but the treatment had no effect on ground diameter. For *M. spectabilis*, the plant height after inoculation increased to varying degrees compared with the ground diameter, but the difference was not significant (**Table 4**). These results of this field experiment support the discovery of multiple PGP mechanisms in the genome of *R. victoriana* JZ-GX1.

DISCUSSION

Plant growth-promoting rhizobacteria have been demonstrated to produce the auxin IAA to enhance plant growth by controlling many physiological processes, such as cell division, vascular tissue differentiation and root elongation (Liu et al., 2019; Luziatelli et al., 2020). However, the biosynthetic pathways of IAA in bacteria are different. Five tryptophan-dependent IAA synthesis pathways have been identified in bacteria, namely, the indole-3-acetonitrile (IAN), tryptamine (TAM), indole-3-acetamide (IAM), indole-3-pyruvate (TPyA), and tryptophan side-chain oxidase (TSO) pathways, and a tryptophan-independent pathway (Zhang et al., 2019b). It has been reported that *R. aquatilis* ZF7 can synthesize IAA through the IPyA pathway (Yuan et al., 2020), and in this study, *R. victoriana* JZ-GX1 was predicted to contain two IAA production pathways (TPyA and IAM). PGPR usually interacts with plant root exudates and synthesizes IAA with the corresponding tryptophan as the precursor (Bhattacharjee et al., 2012; Liu et al., 2016). The JZ-GX1 strain has more than one IAA synthesis pathway, which indicates that it has a stronger ability to secrete IAA than the same microorganism.

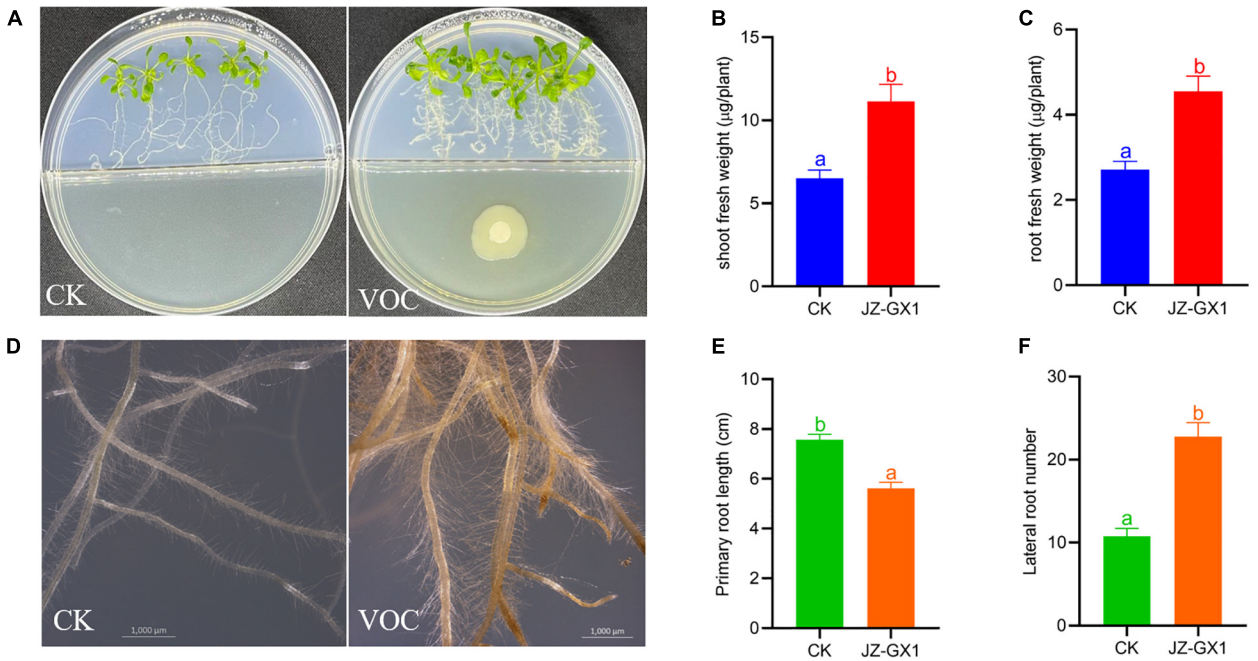


FIGURE 5 | Effect of *Rahnella victoriana* JZ-GX1 VOCs on plant growth and root morphology of *Arabidopsis thaliana* compared to the control (CK). **(A)** Plant phenotype, **(B)** shoot fresh weight, **(C)** root fresh weight, **(D)** root morphology, **(E)** primary root length, **(F)** lateral root length. The scale is 1,000 μm. Different letters indicate statistically significant differences among treatments (Student's *t*-test, $p < 0.01$) ($n = 100$).

Phosphorus (P) is one of the most limiting nutrients required for the growth and development of plants (Zeng et al., 2017; Li et al., 2018). Phytic acid mineralizing rhizobacteria (PMR) play an important role in promoting the dissolution of insoluble organic phosphorus in soil (Kour et al., 2021). *R. victoriana* JZ-GX1 was initially isolated with high phytate-degrading activity as an index, which can promote an increase in the total P content in maize (Li et al., 2013, 2021). However, the mechanism of organophosphorus dissolution has not been clearly revealed. The preliminary determination of the enzymatic properties of *R. victoriana* JZ-GX1 showed that it has a strong ability to hydrolyze phytate under acidic conditions (Li and Wu, 2014). In this study, we found that *R. victoriana* JZ-GX1 harbors two phytase-encoding genes (acid glucose-1-phosphatase and 4-phytase) but lacks appA-like phytase genes. The results of this study are consistent with previously reported enzymatic properties. Until now, there has been very little information available regarding the regulation of phytate-degrading gene expression in bacteria. The presence of phytase genes in this genome could open up opportunities for future molecular cloning and application studies of phytase from *R. victoriana* JZ-GX1.

In the rhizosphere, long-distance interactions through VOCs are an important method of signal transmission between bacteria and plants (Fincheira and Quiroz, 2018; Kong et al., 2021; Li et al., 2021). To date, many bacteria have been shown to stimulate plant growth by releasing volatiles. For example, *B. subtilis* SYST2-derived VOCs can promote tomato growth both *in vitro* and in pot experiments by triggering growth hormone activity (Tahir et al., 2017). Similar findings were obtained for *Arabidopsis* treated with volatiles from *B. megaterium* B55, and the combined effects of blended compounds increased leaf number and leaf surface area up to 2- and 4-fold, respectively, compared with those of the control (Meldau et al., 2012). *B. amyloliquefaciens* GB03 generates 3-hydroxy-2-butanone (acetoin) and 2,3-butanediol as its primary volatiles, which were shown to promote plant growth in trials on *Arabidopsis* (Ryu et al., 2003). The role of these two compounds in plant growth promotion was also confirmed by the addition of the pure acetoin and a mutant bacterium lacking a 2,3-butanediol biosynthesis gene (Farag et al., 2006). In our research, we also detected acetoin in the JZ-GX1 strain by gene prediction and physiological detection, and confirmed that JZ-GX1-derived VOCs could promote an increase in fresh weight and root development without contact with *A. thaliana*. Although the specific mechanism needs to be further explored, this is the first report on the biological activity in plant growth promotion of the VOCs produced by *Rahnella* sp.

When plants are faced with abiotic stress, several important metabolites, including osmotic regulators and active molecules, accumulate in large quantities (Zhou et al., 2017). Spermidine (Spd) is an important plant growth regulator and has also been identified as a protector against various abiotic stressors, such as high salinity, drought, low temperature, and iron deficiency (Zhou et al., 2016a; Li G. et al., 2020; Qian et al., 2021). Reportedly, *B. subtilis* STU6 induces and uses plant arginine to produce more Spd, which promotes bacterial colonization and absorption by the host, thus enhancing the iron acquisition

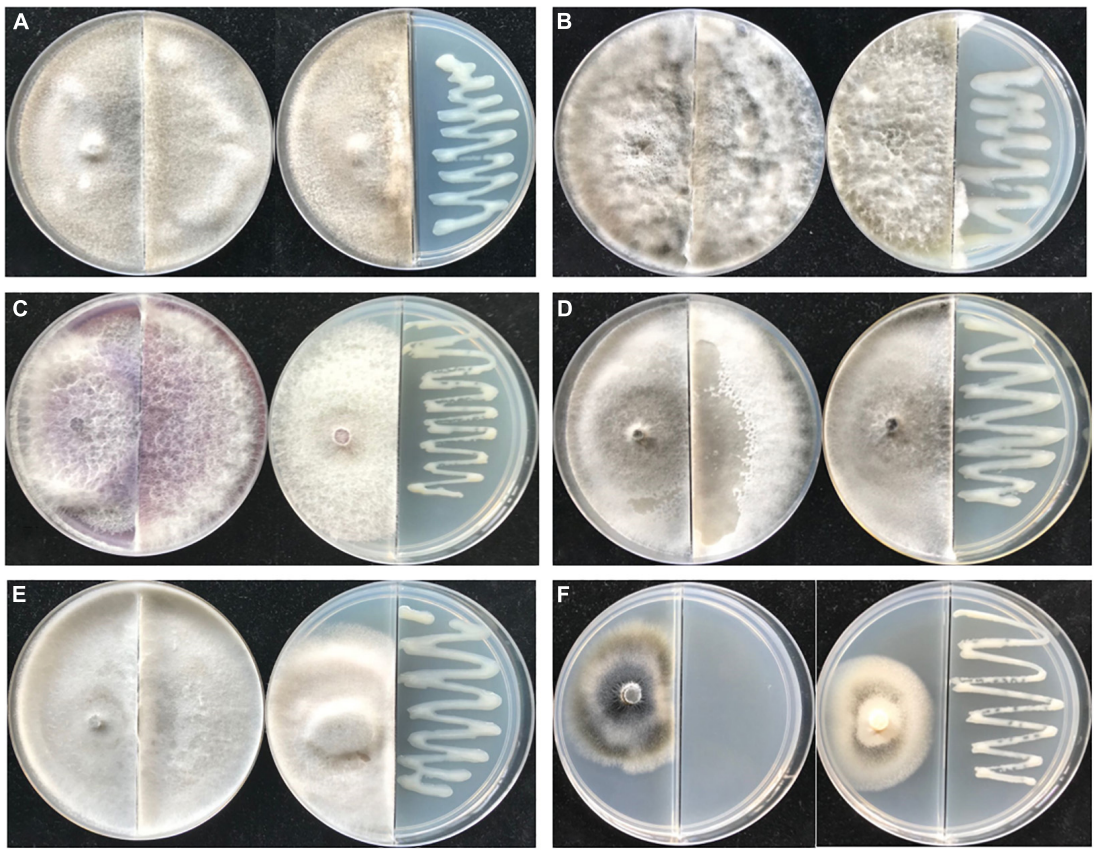


FIGURE 6 | The antifungal spectrum of *Rahnella victoriana* JZ-GX1 VOCs. Plate on the left untreated, plate on the right treated with VOCs. **(A)** *Guignardia camelliae*, **(B)** *Fusicoccus aesculin*, **(C)** *Fusarium oxysporum*, **(D)** *Alternaria tenuissima*, **(E)** *Sphaeropsis sapinea*, and **(F)** *Verticillium dahlia*.

TABLE 4 | Growth-promoting effect of *Rahnella victoriana* JZ-GX1 on different plants.

		Plant height (m)	Ground diameter (mm)
P. euramericana (Dode) Guinier cv. "san Martino" (I-72 poplar)	CK	4.55 ± 0.28a	39.90 ± 4.69a
	JZ-GX1	4.99 ± 0.20b	46.32 ± 3.54b
P. elliottii	CK	49.98 ± 13.49a	13.52 ± 3.79a
	JZ-GX1	61.79 ± 13.42b	15.49 ± 3.12b
P. serotina	CK	147.85 ± 12.99a	17.82 ± 1.70a
	JZ-GX1	168.53 ± 11.05b	17.81 ± 2.98a
M. spectabilis	CK	212.36 ± 7.86a	25.55 ± 3.07a
	JZ-GX1	215.20 ± 10.92a	26.10 ± 3.39a

Means ± standard deviations (n = 90). Different letters indicated a significant difference between control and inoculation treatment based on student's t-test ($p < 0.05$).

system and cell wall iron remobilization in a NO-dependent manner (Zhou et al., 2019). Three genes of spermidine synthase (GE00638, GE00674, and GE00675) were identified in strain JZ-GX1. Furthermore, GABA plays a vital role in plant responses to multiple stressors (Al-Quraan et al., 2013). Guo et al. (2020) found that GABA can improve the tolerance of cucumber to iron deficiency in an auxin-dependent manner. In a previous study, we demonstrated the potential roles of *R. victoriana* JZ-GX1 in alleviating iron deficiency-induced chlorosis in *C. camphora* (Kong et al., 2020b). Here, strain JZ-GX1 was found to possess the genes encoding glutamate decarboxylase, which is involved in GABA production and degradation (GE04524, GE04546, and GE04695). Combined with the previous results, we further clarified the potential mechanism by which *R. victoriana* JZ-GX1 promotes iron uptake in plants. However, the contents of these two substances in plants should be determined, and the genes for the synthesis of these two substances by *R. victoriana* JZ-GX1 should be knocked out to prove this conjecture.

At present, the genus *Rahnella* consists of six closely related species, including *R. aquatilis*, *R. variigena*, *R. victoriana*, *R. bruchi*, *R. woolbedingensis*, and *R. inusitata* (Yuan et al., 2020). Among them, *R. aquatilis* has been widely reported to promote plant growth (Peng et al., 2019; Zhang et al., 2019a; Sun et al., 2020). There are few reports regarding the other species, and so far the only reports of *R. victoriana* was considered associated with decline symptoms on oak and hornbeam (Yousef et al., 2019). In this study, *R. victoriana* JZ-GX1 was proven to be a PGPR strain with a variety of beneficial features, including IAA production, phosphate solubilization, nitrogen fixation, siderophores, acetoin, ACC deaminase, GABA production,

spermidine and VOCs biosynthesis. In addition, the ability of strain JZ-GX1 to promote forest growth has also been well-verified in complex field environments. Overall, the data obtained in this study provide more clues for the potential application of *R. victoriana* JZ-GX1 in the development of eco-friendly biofertilizers, which can promote soil fertility and crop yield.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA720502.

AUTHOR CONTRIBUTIONS

W-LK completed the data analysis and the first draft of the manuscript. W-LK, W-YW, and S-HZ were the finishers of the experimental research. W-YW participated in the experimental result analysis. X-QW directed experimental design, data analysis, and manuscript writing and revision. All authors read and agreed on the final version of manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2017YFD0600104), the Priority Academic Program Development of the Jiangsu Higher Education Institutions (PAPD), and the Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX20_0872).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.828990/full#supplementary-material

REFERENCES

Alexandre, F. S., Della Flora, L. V., Henrique, I. G., Da Silva, D. C., Mercedes, A. P., Silva, A. C., et al. (2021). Arbuscular mycorrhizal fungi (*Rhizophagus clarus*) and rhizobacteria (*Bacillus subtilis*) can improve the clonal propagation and development of teak for commercial plantings. *Front. Plant Sci.* 12:628769. doi: 10.3389/fpls.2021.628769

Al-Quraan, N. A., Sartawe, F. A., and Qaryouti, M. M. (2013). Characterization of gamma-aminobutyric acid metabolism and oxidative damage in wheat (*Triticum aestivum* L.) seedlings under salt and osmotic stress. *J. Plant Physiol.* 170, 1003–1009. doi: 10.1016/j.jplph.2013.02.010

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

Bhattacharjee, R. B., Jourand, P., Chaintreuil, C., Dreyfus, B., Singh, A., and Mukhopadhyay, S. N. (2012). Indole acetic acid and ACC deaminase-producing *Rhizobium leguminosarum* bv. *trifolii* SN10 promote rice growth, and in the process undergo colonization and chemotaxis. *Biol. Fertil. Soils* 48, 173–182. doi: 10.1007/s00374-011-0614-9

Bhattacharyya, D., Garladinne, M., and Lee, Y. H. (2015). Volatile indole produced by rhizobacterium *Proteus vulgaris* JBLS202 stimulates growth of *Arabidopsis thaliana* through auxin, cytokinin, and brassinosteroid pathways. *J. Plant Growth Regul.* 34, 158–168. doi: 10.1007/s00344-014-9453-x

Chandra, P., Wunnava, A., Verma, P., Chandra, A., and Sharma, R. K. (2021). Strategies to mitigate the adverse effect of drought stress on crop plants-influences of soil bacteria: a review. *Pedosphere* 31, 496–509. doi: 10.1016/s1002-0160(20)60092-3

Chandran, M., Manisha, A., and Subashini, A. (2014). Production of phosphate biofertilizer using lignocellulosic waste as carrier material. *Asian J. Chem.* 26, 2065–2067. doi: 10.14233/ajchem.2014.15787

Chen, F., Ye, J., Chio, C., Liu, W., Shi, J., and Qin, W. (2020). A simplified quick microbial genomic DNA extraction via freeze-thawing cycles. *Mol. Biol. Rep.* 47, 703–709. doi: 10.1007/s11033-019-05176-w

Choudhury, A. R., Choi, J., Walitang, D. I., Trivedi, P., Lee, Y., and Sa, T. (2021). ACC deaminase and indole acetic acid producing endophytic bacterial co-inoculation improves physiological traits of red pepper (*Capsicum annum* L.) under salt stress. *J. Plant Physiol.* 267:153544. doi: 10.1016/j.jplph.2021.153544

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Etesami, H., Emami, S., and Alikhani, H. A. (2017). Potassium solubilizing bacteria (KSB): mechanisms, promotion of plant growth, and future prospects - a review. *J. Soil Sci. Plant Nutr.* 17, 897–911. doi: 10.4067/s0718-95162017000400005

Farag, M. A., Ryu, C., Sumner, L. W., and Pare, P. W. (2006). GC-MS SPME profiling of rhizobacterial volatiles reveals prospective inducers of growth promotion and induced systemic resistance in plants. *Phytochemistry* 67, 2262–2268. doi: 10.1016/j.phytochem.2006.07.021

Fincheira, P., and Quiroz, A. (2018). Microbial volatiles as plant growth inducers. *Microbiol. Res.* 208, 63–75. doi: 10.1016/j.micres.2018.01.002

Goswami, M., and Deka, S. (2020). Plant growth-promoting rhizobacteria-alleviators of abiotic stresses in soil: a review. *Pedosphere* 30, 40–61. doi: 10.1016/s1002-0160(19)60839-8

Guo, Z., Du, N., Li, Y., Zheng, S., Shen, S., and Piao, F. (2020). Gamma-aminobutyric acid enhances tolerance to iron deficiency by stimulating auxin signaling in cucumber (*Cucumis sativus* L.). *Ecotoxicol. Environ. Saf.* 192:110285. doi: 10.1016/j.ecoenv.2020.110285

Ha-Tran, D. M., Nguyen, T. T. M., Hung, S., Huang, E., and Huang, C. (2021). Roles of plant growth-promoting rhizobacteria (PGPR) in stimulating salinity stress defense in plants: a review. *Int. J. Mol. Sci.* 22:3154. doi: 10.3390/ijms22063154

Kang, S. M., Asaf, S., Khan, A. L., Khan, A. L., Mun, B. G., Khan, M. A., et al. (2020). Complete genome sequence of *Pseudomonas psychrotolerans* CS51, a plant growth-promoting bacterium, under heavy metal stress conditions. *Microorganisms* 8:382. doi: 10.3390/microorganisms8030382

Kong, W. L., Li, P. S., Wu, X. Q., Wu, T. Y., and Sun, X. R. (2020a). Forest tree associated bacterial diffusible and volatile organic compounds against various phytopathogenic fungi. *Microorganisms* 8:590. doi: 10.3390/microorganisms8040590

Kong, W. L., Wu, X. Q., and Zhao, Y. J. (2020b). Effects of *Rahnella aquatilis* JZ-GX1 on treat chlorosis induced by iron deficiency in *Cinnamomum camphora*. *J. Plant Growth Regul.* 39, 877–887. doi: 10.1007/s00344-019-10029-8

Kong, W. L., Wang, Y. H., and Wu, X. Q. (2021). Enhanced iron uptake in plants by volatile emissions of *Rahnella aquatilis* JZ-GX1. *Front. Plant Sci.* 12:704000. doi: 10.3389/fpls.2021.704000

Kour, D., Rana, K. L., Kaur, T., Yadav, N., Yadav, A. N., Kumar, M., et al. (2021). Biodiversity, current developments and potential biotechnological applications of phosphorus-solubilizing and -mobilizing microbes: a review. *Pedosphere* 31, 43–75. doi: 10.1016/s1002-0160(20)60057-1

Kusale, S. P., Attar, Y. C., Sayyed, R. Z., Malek, R. A., Ilyas, N., Suriani, N. L., et al. (2021). Production of plant beneficial and antioxidants metabolites by *Klebsiella variicola* under salinity stress. *Molecules* 26:1894. doi: 10.3390/molecules26071894

Li, F., Shi, T., Tang, X., Tang, M., Gong, J., and Yi, Y. (2020). *Bacillus amyloliquefaciens* PDR1 from root of karst adaptive plant enhances *Arabidopsis thaliana* resistance to alkaline stress through modulation of plasma membrane H^+-ATPase activity. *Plant Physiol. Biochem.* 155, 472–482. doi: 10.1016/j.plaphy.2020.08.019

Li, G., Liang, Z., Li, Y., Liao, Y., and Liu, Y. (2020b). Exogenous spermidine regulates starch synthesis and the antioxidant system to promote wheat grain filling under drought stress. *Acta Physiol. Plant.* 42, 1723–1734.

Li, G., Wu, X., Ye, J., Hou, L., Zhou, A., and Zhao, L. (2013). Isolation and identification of phytate-degrading rhizobacteria with activity of improving growth of poplar and Masson pine. *World J. Microbiol. Biotechnol.* 29, 2181–2193. doi: 10.1007/s11274-013-1384-3

Li, G. E., and Wu, X. Q. (2014). Study on enzymology characteristics of phytase secreted by efficient phytate-degrading rhizobacteria *Rahnella aquatilis* JZ-GX1 (in Chinese). *J. Cent. S. Univ. For. Technol.* 34, 90–93.

Li, G. X., Wu, X. Q., Ye, J. R., and Yang, H. (2018). Characteristics of organic acid secretion associated with the interaction between *Burkholderia multivorans* WS-FJ9 and poplar root system. *Biomed Res. Int.* 2018:9619724. doi: 10.1155/2018/9619724

Li, P. S., Kong, W. L., Wu, X. Q., and Zhang, Y. (2021). Volatile organic compounds of the plant growth-promoting rhizobacteria JZ-GX1 enhanced the tolerance of *Robinia pseudoacacia* to salt stress. *Front. Plant Sci.* 12:753332. doi: 10.3389/fpls.2021.753332

Liu, A., Zhang, P., Bai, B., Bai, F., Jin, T., and Ren, J. (2020a). Volatile organic compounds of endophytic *Burkholderia pyrrocinia* strain JK-SH007 promote disease resistance in poplar. *Plant Dis.* 104, 1610–1620. doi: 10.1094/PDIS-11-19-2366-RE

Liu, S., Tian, Y., Jia, M., Lu, X., Yue, L., Zhao, X., et al. (2020b). Induction of salt tolerance in *Arabidopsis thaliana* by volatiles from *Bacillus amyloliquefaciens* FZB42 via the jasmonic acid signaling pathway. *Front. Microbiol.* 11:562934. doi: 10.3389/fmicb.2020.562934

Liu, Y., Wang, Y., Kong, W., Liu, W., Xie, X., and Wu, X. (2020c). Identification, cloning and expression patterns of the genes related to phosphate solubilization in *Burkholderia multivorans* WS-FJ9 under different soluble phosphate levels. *AMB Express* 10:108. doi: 10.1186/s13568-020-01032-4

Liu, W., Chen, F., Wang, C., Fu, H., Fang, X., Ye, J., et al. (2019). Indole-3-acetic acid in *Burkholderia pyrrocinia* JK-SH007: enzymatic identification of the indole-3-acetamide synthesis pathway. *Front. Microbiol.* 10:2559. doi: 10.3389/fmicb.2019.02559

Liu, X., Li, X., Li, Y., Li, R., and Xie, Z. (2017). Plant growth promotion properties of bacterial strains isolated from the rhizosphere of the Jerusalem artichoke (*Helianthus tuberosus* L.) adapted to saline-alkaline soils and their effect on wheat growth. *Can. J. Microbiol.* 63, 228–237. doi: 10.1139/cjm-2016-0511

Liu, Y., Chen, L., Zhang, N., Li, Z., Zhang, G., Xu, Y., et al. (2016). Plant-microbe communication enhances auxin biosynthesis by a root-associated bacterium, *Bacillus amyloliquefaciens* SQR9. *Mol. Plant Microbe Interact.* 29, 324–330. doi: 10.1094/MPMI-10-15-0239-R

Luziatelli, F., Ficca, A. G., Cardarelli, M., Melini, F., Cavalieri, A., and Ruzzi, M. (2020). Genome sequencing of *Pantoea agglomerans* C1 provides insights into molecular and genetic mechanisms of plant growth-promotion and tolerance to heavy metals. *Microorganisms* 8:153. doi: 10.3390/microorganisms8020153

Mariotti, L., Scartazza, A., Curadi, M., Picciarelli, P., and Toffanin, A. (2021). *Azospirillum baldaniorum* Sp245 induces physiological responses to alleviate the adverse effects of drought stress in purple basil. *Plants* 10:1141. doi: 10.3390/plants10061141

Meldau, D. G., Long, H. H., and Baldwin, I. T. (2012). A native plant growth promoting bacterium, *Bacillus* sp. B55, rescues growth performance of an ethylene-insensitive plant genotype in nature. *Front. Plant Sci.* 3:112. doi: 10.3389/fpls.2012.00112

Mondal, S., Pramanik, K., Ghosh, S. K., Pal, P., Mondal, T., Soren, T., et al. (2021). Unraveling the role of plant growth-promoting rhizobacteria in the alleviation of arsenic phytotoxicity: a review. *Microbiol. Res.* 250:126809. doi: 10.1016/j.micres.2021.126809

Nascimento, F. X., Hernandez, A. G., Glick, B. R., and Rossi, M. J. (2020). Plant growth-promoting activities and genomic analysis of the stress-resistant *Bacillus megaterium* STB1, a bacterium of agricultural and biotechnological interest. *Biotechnol. Rep.* 25:e406. doi: 10.1016/j.btre.2019.e00406

Olenska, E., Malek, W., Wojcik, I., Swiecicka, I., Thijs, S., and Vangronsveld, J. (2020). Beneficial features of plant growth-promoting rhizobacteria for improving plant growth and health in challenging conditions: a methodical review. *Sci. Total Environ.* 743:140682–140682. doi: 10.1016/j.scitotenv.2020.140682

Peng, J., Wu, D., Liang, Y., and Guo, Y. B. (2019). Disruption of acdS gene reduces plant growth promotion activity and maize saline stress resistance by *Rahnella aquatilis* HX2. *J. Basic Microbiol.* 59, 402–411. doi: 10.1002/jobm.201800510

Penrose, D. M., and Glick, B. R. (2003). Methods for isolating and characterizing ACC deaminase-containing PGPR. *Physiol. Plant.* 118, 10–15. doi: 10.1034/j.1399-3054.2003.00086.x

Perez-Flores, P., Valencia-Cantero, E., Altamirano-Hernandez, J., Pelagio-Flores, R., Lopez-Bucio, J., Garcia-Juarez, P., et al. (2017). *Bacillus methylotrophicus* M4-96 isolated from maize (*Zea mays*) rhizoplane increases growth and auxin content in *Arabidopsis thaliana* via emission of volatiles. *Protoplasma* 254, 2201–2213. doi: 10.1007/s00709-017-1109-9

Pretty, J., Benton, T. G., Bharucha, Z. P., Dicks, L. C., Flora, C. B., Godfray, H. C. J., et al. (2018). Global assessment of agricultural system redesign for sustainable intensification. *Nat. Sustain.* 1, 441–446. doi: 10.1038/s41893-018-0114-0

Qian, R., Ma, X., Zhang, X., Hu, Q., Liu, H., and Zheng, J. (2021). Effect of exogenous spermidine on osmotic adjustment, antioxidant enzymes activity, and gene expression of *Gladiolus gandavensis* seedlings under salt stress. *J. Plant Growth Regul.* 40, 1353–1367. doi: 10.1007/s00344-020-10198-x

Rath, M., Mitchell, T. R., and Gold, S. E. (2018). Volatiles produced by *Bacillus mojavensis* RRC101 act as plant growth modulators and are strongly culture-dependent. *Microbiol. Res.* 208, 76–84. doi: 10.1016/j.micres.2017.12.014

Rosier, A., Medeiros, F. H. V., and Bais, H. P. (2018). Defining plant growth promoting rhizobacteria molecular and biochemical networks in beneficial plant-microbe interactions. *Plant Soil* 428, 35–55. doi: 10.1007/s11104-018-3679-5

Ruzzi, M., and Aroca, R. (2015). Plant growth-promoting rhizobacteria act as biostimulants in horticulture. *Sci. Hortic.* 196, 124–134. doi: 10.1016/j.scienta.2015.08.042

Ryu, C. M., Farag, M. A., Hu, C. H., Reddy, M. S., Wei, H. X., Pare, P. W., et al. (2003). Bacterial volatiles promote growth in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4927–4932. doi: 10.1073/pnas.0730845100

Schwyn, B., and Neilands, J. B. (1987). Universal chemical assay for the detection and determination of siderophores. *Anal. Biochem.* 160, 47–56. doi: 10.1016/0003-2697(87)90612-9

Shen, L., Wu, X. Q., Zeng, Q. W., and Liu, H. (2016). Regulation of soluble phosphate on the ability of phytate mineralization and beta-propeller phytase gene expression of *Pseudomonas fluorescens* JZ-DZ1, a phytate-mineralizing rhizobacterium. *Curr. Microbiol.* 73, 915–923. doi: 10.1007/s00284-016-1139-0

Sun, X. W., Ma, W., Xu, Y., Jing, X. M., and Ni, H. (2020). Complete genome sequence of *Rahnella aquatilis* MEM40, a plant growth-promoting rhizobacterium isolated from rice rhizosphere soil, with antagonism against *Magnaporthe oryzae* and *Fusarium graminearum*. *Microbiol. Resour. Announc.* 9:e00651-20. doi: 10.1128/MRA.00651-20

Tahir, H. A. S., Gu, Q., Wu, H., Raza, W., Hanif, A., Wu, L., et al. (2017). Plant growth promotion by volatile organic compounds produced by *Bacillus subtilis* SYST2. *Front. Microbiol.* 8:171. doi: 10.3389/fmicb.2017.00171

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41

Teshome, D. T., Zharare, G. E., and Naidoo, S. (2020). The threat of the combined effect of biotic and abiotic stress factors in forestry under a changing climate. *Front. Plant Sci.* 11:601009. doi: 10.3389/fpls.2020.601009

Wang, Y., Hou, L., Wu, X., Zhu, M., Dai, Y., and Zhao, Y. (2021). Mycorrhiza helper bacterium *Bacillus pumilus* HR10 the improves growth and nutritional status of *Pinus thunbergii* by promoting mycorrhizal proliferation. *Tree Physiol.* tab139. doi: 10.1093/treephys/tpab139

Yasmin, H., Bano, A., Wilson, N. L., Nosheen, A., Naz, R., Hassan, M. N., et al. (2021). Drought-tolerant *Pseudomonas* sp. showed differential expression of stress-responsive genes and induced drought tolerance in *Arabidopsis thaliana*. *Physiol. Plant.* 174:e13497. doi: 10.1111/ppl.13497

Yousef, M., Heshmat, R., Valiollah, B., and Sandra, D. (2019). *Brenneria* spp. and *Rahnella victoriana* associated with acute oak decline symptoms on oak and hornbeam in Iran. *For. Pathol.* 49:e12535. doi: 10.1111/efp.12535

Yuan, L., Li, L., Zheng, F., Shi, Y., Xie, X., Chai, A., et al. (2020). The complete genome sequence of *Rahnella aquatilis* ZF7 reveals potential beneficial properties and stress tolerance capabilities. *Arch. Microbiol.* 202, 483–499. doi: 10.1007/s00203-019-01758-1

Zeng, Q., Wu, X., Wang, J., and Ding, X. (2017). Phosphate solubilization and gene expression of phosphate-solubilizing bacterium *Burkholderia multivorans* WS-FJ9 under different levels of soluble phosphate. *J. Microbiol. Biotechnol.* 27, 844–855. doi: 10.4014/jmb.1611.11057

Zhang, L., Fan, J. Q., Feng, G., and Declerck, S. (2019a). The arbuscular mycorrhizal fungus *Rhizophagus irregularis* MUCL 43194 induces the gene expression of citrate synthase in the tricarboxylic acid cycle of the phosphate-solubilizing bacterium *Rahnella aquatilis* HX2. *Mycorrhiza* 29, 69–75.

Zhang, P., Jin, T., Sahu, S. K., Xu, J., Shi, Q., Liu, H., et al. (2019b). The distribution of tryptophan-dependent indole-3-acetic acid synthesis pathways in bacteria unraveled by large-scale genomic analysis. *Molecules* 24:1411. doi: 10.3390/molecules24071411

Zhang, Y., Li, T., Liu, Y., Li, X., Zhang, C., Feng, Z., et al. (2019c). Volatile organic compounds produced by *Pseudomonas chlororaphis* subsp. *aureofaciens* SPS-41 as biological fumigants to control *Ceratocystis fimbriata* in postharvest sweet potatoes. *J. Agric. Food Chem.* 67, 3702–3710. doi: 10.1021/acs.jafc.9b00289

Zhang, Y., Li, T., Xu, M., Guo, J., Zhang, C., Feng, Z., et al. (2021). Antifungal effect of volatile organic compounds produced by *Pseudomonas chlororaphis* subsp. *aureofaciens* SPS-41 on oxidative stress and mitochondrial dysfunction of *Ceratocystis fimbriata*. *Pestic. Biochem. Physiol.* 173:104777. doi: 10.1016/j.pestbp.2021.104777

Zhou, C., Liu, Z., Zhu, L., Ma, Z., Wang, J., and Zhu, J. (2016a). Exogenous melatonin improves plant iron deficiency tolerance via increased accumulation of polyamine-mediated nitric oxide. *Int. J. Mol. Sci.* 17:1777. doi: 10.3390/ijms17111777

Zhou, C., Ma, Z., Zhu, L., Xiao, X., Xie, Y., Zhu, J., et al. (2016b). Rhizobacterial strain *Bacillus megaterium* BOFC15 induces cellular polyamine changes that improve plant growth and drought resistance. *Int. J. Mol. Sci.* 17:976. doi: 10.3390/ijms17060976

Zhou, C., Zhu, L., Guo, J., Xiao, X., Ma, Z., and Wang, J. (2019). *Bacillus subtilis* STU6 ameliorates iron deficiency in tomato by enhancement of polyamine-mediated iron remobilization. *J. Agric. Food Chem.* 67, 320–330. doi: 10.1021/acs.jafc.8b05851

Zhou, C., Zhu, L., Xie, Y., Li, F., Xiao, X., Ma, Z., et al. (2017). *Bacillus licheniformis* SA03 confers increased saline-alkaline tolerance in *Chrysanthemum* plants by induction of abscisic acid accumulation. *Front. Plant Sci.* 8:1143. doi: 10.3389/fpls.2017.01143

Conflict of Interest: 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fmicb.2022.862776

Pan-Genome Analysis of *Laribacter hongkongensis*: Virulence Gene Profiles, Carbohydrate-Active Enzyme Prediction, and Antimicrobial Resistance Characterization

*Pei-Bo Yuan†, Yi Zhan†, Jia-Hui Zhu, Jia-Hui Ling, En-Zhong Chen, Wan-Ting Liu, Lin-Jing Wang, Yu-Xia Zhong and Ding-Qiang Chen**

Department of Laboratory Medicine, Microbiome Medicine Center, Zhujiang Hospital, Southern Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Feng Gao,
Tianjin University, China

Reviewed by:
Yongbing Zhao,
National Institutes of Health (NIH), United States
Hualin Liu,
Sun Yat-sen University, China

***Correspondence:**
Ding-Qiang Chen
jyksys@126.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: *26 January 2022*
Accepted: *08 March 2022*
Published: *31 March 2022*

Citation:
Yuan P-B, Zhan Y, Zhu J-H, Ling J-H, Chen E-Z, Liu W-T, Wang L-J, Zhong Y-X and Chen D-Q (2022) Pan-Genome Analysis of Laribacter hongkongensis: Virulence Gene Profiles, Carbohydrate-Active Enzyme Prediction, and Antimicrobial Resistance Characterization. Front. Microbiol. 13:862776. doi: 10.3389/fmicb.2022.862776

Laribacter hongkongensis is a new emerging foodborne pathogen that causes community-acquired gastroenteritis and traveler's diarrhea. However, the genetic features of *L. hongkongensis* have not yet been properly understood. A total of 45 aquatic animal-associated *L. hongkongensis* strains isolated from intestinal specimens of frogs and grass carps were subjected to whole-genome sequencing (WGS), along with the genome data of 4 reported human clinical strains, the analysis of virulence genes, carbohydrate-active enzymes, and antimicrobial resistance (AMR) determinants were carried out for comprehensively understanding of this new foodborne pathogen. Human clinical strains were genetically more related to some strains from frogs inferred from phylogenetic trees. The distribution of virulence genes and carbohydrate-active enzymes exhibited different patterns among strains of different sources, reflecting their adaption to different host environments and indicating different potentials to infect humans. Thirty-two AMR genes were detected, susceptibility to 18 clinical used antibiotics including aminoglycoside, chloramphenicol, trimethoprim, and sulfa was checked to evaluate the availability of clinical medicines. Resistance to Rifampicin, Cefazolin, ceftazidime, Ampicillin, and ceftriaxone is prevalent in most strains, resistance to tetracycline, trimethoprim-sulfamethoxazole, ciprofloxacin, and levofloxacin are aggregated in nearly half of frog-derived strains, suggesting that drug resistance of frog-derived strains is more serious, and clinical treatment for *L. hongkongensis* infection should be more cautious.

Keywords: pan-genome analysis, *L. hongkongensis*, virulence gene, carbohydrate-active enzyme, antimicrobial resistance

INTRODUCTION

Laribacter hongkongensis (*L. hongkongensis*) is a gram-negative foodborne organism. It is firstly discovered in the stool of six patients with diarrhea in Hong Kong and then reported to be associated with community-acquired gastroenteritis and stomach and intestinal infection or traveler's diarrhea (Yuen et al., 2001; Woo et al., 2004), and is suggested to be distributed around

the world according to the global case reports (Teng et al., 2005; Woo et al., 2005; Kim et al., 2011; Beilfuss et al., 2015; Engsbro et al., 2018; Hung et al., 2020). *L. hongkongensis* has been also discovered in the gut of diverse aquatic animals, birds, drinking water reservoirs, aquatic environments, as well as wastewaters (Lau et al., 2007a,b, 2009; Ni et al., 2011; Greay et al., 2019). Frog and freshwater fish have been recognized to be the primary hosts for *L. hongkongensis* and the sources of human infection. Eating raw fish, as well as inadequate food cleaning and disinfection, can lead to an increase in foodborne *L. hongkongensis* infections.

Since *L. hongkongensis* was identified and confirmed as pathogenic by Koch's Law (Beilfuss et al., 2015), a series of studies have been conducted to understand its pathogenicity and epidemiology. The serological detection methods were established (Tse et al., 2014; Wang et al., 2017). The adaptive response and transcriptional regulation patterns of *L. hongkongensis* coupled with different nutritional sources, aerobic or anaerobic conditions, and different temperatures were also studied (Xiong L. et al., 2015; Kong et al., 2016, 2017; Xiong et al., 2017, 2019). Researchers also discussed the pathogenicity factors of *L. hongkongensis* (Xie et al., 2014). Drug-resistant properties and the presence of clinical multiple drug-resistant strains have also been reported (Raja and Ghosh, 2014; Wu et al., 2018; Teng et al., 2021). However, these studies were conducted on specific isolated strains. And all 4 genomes published were isolated from clinical human samples. To better understand the genetic and pathogenic features of *L. hongkongensis*, and to assess the risk of infection from the main primary hosts of *L. hongkongensis*, a large-scale genome-based analysis involving strains from Frogs and freshwater fishes should be carried out.

Here, we conducted a pan-genome analysis of 45 aquatic frogs and grass carps isolated strains and all 4 clinically reported strains with genome data. A comprehensive analysis of the genetic evolution, virulence factors, carbohydrate-active enzymes, drug-resistant genes, and phenotypes were performed to assess the influence of bacteria's source and lineage on pathogenicity risk, providing a reference for further studies on source tracing and treatment of *L. hongkongensis* related infections.

MATERIALS AND METHODS

Bacterial Isolation

Intestinal specimens of grass carp and frog were collected from retail markets in Guangzhou. Through isolation, culture, biochemical identification, 45 strains were identified as *L. hongkongensis*. Among the 45 strains, 21 strains were isolated from grass carps, and 24 strains from frogs. Strains isolated from frog were named with an initial letter W (indicating WA in Chinese), and strains isolated from grass carp were named with an initial letter Y (indicating YU in Chinese).

Preparation of Genomic DNA and Whole-Genome Sequencing

Genomic DNA was extracted from freshly grown cells by a bacterial DNA extraction kit (Dongsheng Biotech, CHN). The genomic DNA was mechanically sheared using a Nebulizer instrument (Invitrogen) to select fragments of approximately 550 bp. A DNA library was prepared using the Illumina TruSeq^{TM} Nano method and sequenced on the Illumina MiSeq platform with the 2 × 250 bp paired-end (PE) reagent kit v2. The quality of the raw reads was checked in Fast QC v0.11.5,[1] and low-quality reads were filtered. Reads with adapter, ≥ 10% unknown bases, or > 40% low-quality bases (Q ≤ 10) were designated as low-quality reads. Filtered clean reads were assembled into contigs with edena v3 by default parameters (Hernandez et al., 2014), which is based on the classical graph approach and known for its efficiency in handling base errors and detecting potentially spurious reads. The genome assembly quality was further evaluated by quast v5.0.2 (Gurevich et al., 2013).

The genome sequences and associated data for 45 new sequenced *L. hongkongensis* strains reported in this study were deposited in NCBI under the BioProject accession number PRJNA770832.

Genome Preparation, Annotation, and Pan-Genome Inference

Four *L. hongkongensis* genome sequences of human clinical strains were downloaded from NCBI.[2] All 49 *L. hongkongensis* genomes were annotated *de novo* with Prokka (Seemann, 2014) version 1.14.5. Function annotation and orthology prediction of all genes was conducted using eggNOG 5.0 (Huerta-Cepas et al., 2019). The Prokka produced GFF3 format annotation files were provided to the rapid large-scale prokaryote pan-genome analysis tool Roary version 3.12.0 (Page et al., 2015) to calculate the pan-genome of *L. hongkongensis*. Coding regions were extracted from the input and converted to protein sequences, filtered to remove partial sequences, and iteratively pre-clustered with CD-HIT, then an all-against-all comparison was performed with a built-in BLASTP on the reduced sequences with the default sequence identity cutoff of 95%. Roary produced a gene presence/absence matrix, a multi-FASTA alignment of core genes using PRANK (31) version 170427, and determined the subsets genes of core and pan-genomes. The gene presence/absence matrix was visualized using the built-in heatmap function of R with presence shown in red and absence in blue. The strains were clustered by row-side dendrogram generated with the default complete linkage clustering method using the euclidean distance measure.

For the statistical extrapolation, non-linear least squares curve fittings of the observed core and pan genome sizes as functions of the number of analyzed genomes were performed with GraphPad Prism 8 following the models/regression algorithms given by researchers (Tettelin et al., 2005, 2008; Nourdin-Galindo et al., 2017). Curve fitting the pan-genome was performed using a power-law regression based on Heaps' law [$y = A_{pan}x^{B_{pan}}$], as previously described (Tettelin et al., 2008; Nourdin-Galindo et al., 2017), where y was the pan-genome size, x was the number of analyzed genomes, A_{pan} and B_{pan} were the fitting parameters. B_{pan} was equivalent to the γ parameter in estimating an open or closed pan-genome (Tettelin et al., 2008). When

[1] https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[2] https://www.ncbi.nlm.nih.gov/

$0 < B_{pan} < 1$ indicates an open pan-genome, while $B_{pan} < 0$ or $B_{pan} > 1$ indicate a closed pan-genome. Curve fitting of the core-genome was performed using an exponential decay regression model [$y = A_{core}e^{(-B_{core}\,x)}C_{core}$], where y was the core-genome size, x was the number of analyzed genomes, C_{core} was the extrapolated size of the core genome for $x\rightarrow\infty$, A_{core} and B_{core} were the fitting parameters. Curve fittings of new genes discovered with the availability of additional genome sequences were performed using both a power-law regression based on Heaps' law [$y = A_{new}x^{-B_{new}}$], and an exponential decay regression model [$y = C_{new}e^{(-D_{new}\,x)}E_{new}$], where y was the new gene number, x was the number of analyzed genomes, and A_{new}, B_{new}, C_{new}, D_{new}, and E_{new} were the fitting parameters. The B_{new} was equivalent to the α parameter in estimating an open or closed pan-genome ($\alpha \leq 1$, the pan-genome is open; $\alpha > 1$, the pan-genome is closed), while E_{new} was equivalent to the tg(θ) parameter representing the number of new genes asymptotically predicted for further genome sequencing (Tettelin et al., 2008).

Phylogenetic Analyses Based on Core Genome

A phylogenetic tree of 49 *L. hongkongensis* strains was built based on protein sequences of core genes with the PhyloPhlAn pipeline v3.0.60 (Asnicar et al., 2020). Core genes were determined by Roary, and their corresponding protein sequences from HKU1 were extracted as markers. Amino acid sequences predicted by prokka for core genes were extracted from genomes as inputs. The programs FastTree and RAxML (Stamatakis, 2014) were used to build the trees using the PhyloPhlAn 3.0.60 database in an accurate mode with the diversity level set as low. Specifically, homologs were first identified and extracted using Diamond 0.9.21 (Buchfink et al., 2021) with the following command "blastp –more-sensitive –id 50 –max-hsps 35 -k 0." Each variant of each marker was then aligned using MAFFT 7.487 (Katoh and Standley, 2013), with command "–anysymbol –auto." The generated multiple-sequence alignments were concatenated and trimmed using trimAl 1.4.rev15 (Capella-Gutiérrez et al., 2009), with "-gappyout" option. A maximum likelihood phylogenetic tree was produced with LG-CAT model using FastTree 2.1.10 (Price et al., 2010), with command "-quiet -pseudo -spr 4 -mlacc 2 -slownni -fastest -no2nd -mlnni 4 –lg." The phylogeny was then refined using RAxML 8.2.12 (Stamatakis, 2014), with "-p 1989 -m PROTCATLG" options. The tree was rooted by the mid-point method, and visualized and edited online at the website of iTOL (Letunic and Bork, 2021).

Virulence Factors Identification

Prediction of virulence genes was performed through searches querying each genome against the Virulence Factor Database (Liu et al., 2019) with the VFanalyzer web server. All parameters used were set as default. Based on sequences of Virulence Factor, another phylogenetic tree of 49 *L. hongkongensis* strains was built with Parsnp 1.5.6 (Treangen et al., 2014), then rooted by the mid-point method and visualized with iTOL v5 (Letunic and Bork, 2021).

Carbohydrate-Active enZYme Annotation

The fasta file of 49 *L. hongkongensis* strains was subjected to automated annotation and assignment to CAZymes using the dbCAN2 meta server (Zhang et al., 2018). Three tools, HMMER (for annotated CAZyme domain boundaries according to the dbCAN CAZyme domain HMM database), DIAMOND (for fast blast hits in the CAZy database), and Hoptep (for short conserved motifs in the PPR library) were used with the default *E*-value and coverage cutoff. All descriptions and classifications were compiled from CAZy (Lombard et al., 2014).

Antimicrobial Resistance Determinants Identification and Antimicrobial Susceptibility Testing

Prediction of antibiotic resistance genes was performed using RGI (v.3.1.1) (Alcock et al., 2020). All isolates were tested for antimicrobial susceptibility by the disk diffusion method using Mueller-Hinton agar and commercially available discs (Oxoid). The antimicrobial agents used were Amikacin (30 μg), Ampicillin (10 μg), Ampicillin-Sulbactam (30 μg), Aztreonam (30 μg), Cefazolin (30 μg), Cefepime (10 μg), Ceftazidime (30 μg), Ceftriaxone (30 μg), Cefuroxime (30 μg), Ciprofloxacin (5 μg), Gentamicin (10 μg), Imipenem (10 μg), Levofloxacin (5 μg), Meropenem (10 μg), Minocycline (30 μg), Rifampicin (5 μg), Tetracycline (30 μg), Trimethoprim-Sulfamethoxazole (25 μg). *Escherichia coli* ATCC25922 is used as a quality control organism. The Antibiotic susceptibility was interpreted based on Clinical and Laboratory Standards Institution (CLSI) guidelines.

RESULTS AND DISCUSSION

Pan-Genome Overview

The average full genome size and GC content of the 49 *L. hongkongensis* strains were 3.19 Mb and 62.5%, with an average of 2,997 coding sequences, respectively. The general characteristics of *L. hongkongensis* genomes can be found in **Supplementary Table 1**. The pan-genome of *L. hongkongensis* was inferred with Roary. Roary produced a total of 8,964 protein-coding gene sequence clusters. The "core genome," consisting of genes present in at least 95% of strains (≥47 out of 49), was represented by 2,291 genes (≈26% of all genes). The 6,673 non-core genes were divided into 1,341 "shell genes" (i.e., shell genes present in at least 15% of strains, and no more than 95% of strains; ≈15% of all genes), and 5,332 "cloud genes" (i.e., cloud genes present in no more than 15% strains, ≤ 7 out of 49 strains; ≈59% of all genes) (**Figure 1A**). A heatmap was drawn to visualize the presence/absence of all 8,964 genes in 49 genomes (**Figure 1B**). The strains can be divided into 6 clusters for holding specific genes. Interestingly, there were 203 genes only held by some strains from grass carps (the second cluster in brown word label). Though most of these genes can only be annotated as function poorly characterized, some genes were annotated to be involved with information storage and processing, indicating they might be important for their adaptation of intestinal specimens of

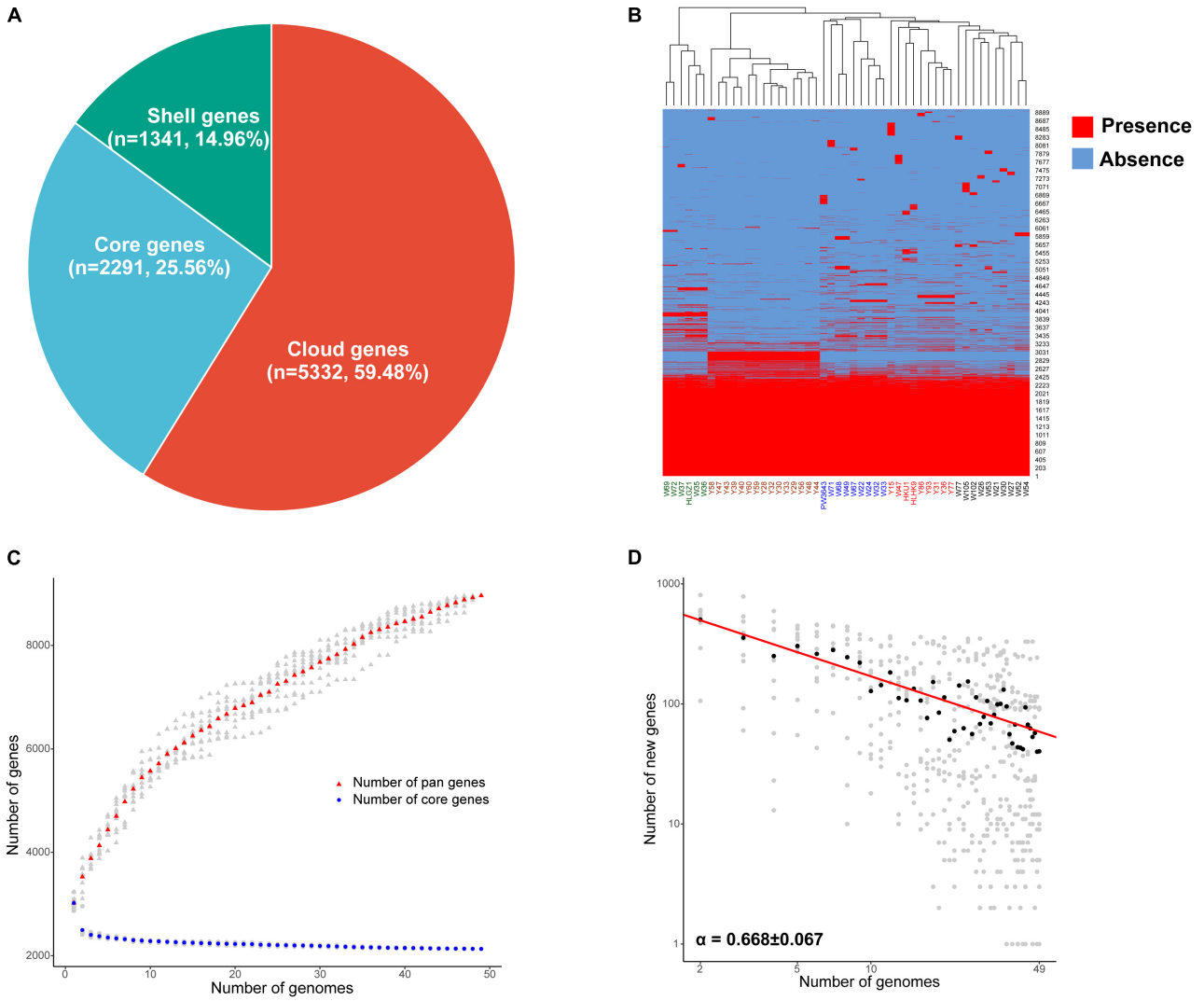


FIGURE 1 | Pan-genome analysis of 49 *L. hongkongensis* strains. **(A)** Composition of core, shell, and cloud genes in pan-genome. **(B)** Presence and absence of all identified genes. Presence was shown in red and absence was shown in blue, the clustering of samples based on genes' presence and absence was displayed above the heatmap. **(C)** The core and pan-genome development plots. The number of pan genes was shown by red dots, and core genes by blue dots, with *L. hongkongensis* genome sizes shown on the horizontal axis. **(D)** The log-log plot of Power law regression for new genes discovered with the availability of additional genome sequences. Medians of the distributions were indicated by black dots, the exponent α was from the Heaps' law $n = kN^{-\alpha}$, where n was the new gene number, N was the number of analyzed genomes, α can be used to estimating an open or closed pan-genome ($\alpha \leq 1$, the pan-genome is open; $\alpha > 1$, the pan-genome is closed). Strains isolated from frog were named with an initial letter W, and strains isolated from grass carp were named with an initial letter Y, human clinical strains were presented in strong pink, strains named HKU1, HLHK9, and PW3643 were isolated from Hong Kong patients, while HLGZ1 was isolated from a Guangzhou patient.

grass carps. The detailed gene presence and absence matrix and annotation of all genes can be found in **Supplementary Table 2**.

To determine whether the pan-genome was open or closed, the core and pan-genome development plots of *L. hongkongensis* were drawn as **Figure 1C**, the number of genes in the pan-genome increased as more genomes were sequenced, whereas the number of genes in the core-genome decreased. The pan-genome of *L. hongkongensis* could be considered "open," as supported by the growth exponent value of 0.3019 ± 0.005 (95% confidence interval), based on B_{pan} (equivalent to γ in the reference research) parameter from Heaps' law. The exponential decay regression fitting of core-genome revealed an extrapolated core genome size of 2,198 (95% confidence interval 2,192–2,205), based on the value of the C_{core} parameter. Furthermore, the power law and exponential regression fits for new genes discovered with the availability of additional genome sequences were performed, and the log-log plot was built (**Figure 1D**). The B_{new} (equivalent to α in the reference research) parameter of power law ($B_{new} = 0.668 \pm 0.066 < 1$) was in good agreement with the value of B_{pan} ($0 < B_{pan} = 0.3019 < 1$), indicating the openness of the pan-genome. The value of parameter E_{new} [tg(θ)] in exponential regression fit indicated that each additional

genome sequence would add asymptotic 71.24 ± 14.91 (95% confidence interval) new genes to the pan-genome. The open pan-genome suggested the ability of *L. hongkongensis* to adapt to new niches by generating or incorporating new genes, which might support its virulence and host adaption.

Function Annotation and Phylogenetic Analysis of Core Genes

The functions of 2,291 core genes were characterized by matching the sequences with the COG database, giving out the conserved functions which may play housekeeping roles in *L. hongkongensis* (**Figure 2A** and **Supplementary Table 3**). The core genes were annotated in metabolism (37.54%), information storage and processing (16.32%), cellular processes and signaling (19.03%), and about 28% of core genes are functions unknown. It is worth mentioning that numerous core genes are annotated in processes like signal transduction mechanisms (87 genes), intracellular trafficking, secretion, and vesicular transport (32 genes), defense mechanism (27 genes), indicating that these processes may be important for the survival of *L. hongkongensis* in the various living environments and may play key roles in interaction with host immune systems.

A phylogenetic tree of 49 *L. hongkongensis* strains was constructed based on amino acid sequences of 2,291 core genes to understand the genetic relationship between strains from different sources (**Figure 2B**). The strains can be divided into four clusters in the tree. The strains in the core I cluster and core II cluster were derived from the grass carp samples, while strains in the core III and core IV were mostly derived from the frog samples except for HLGZ1, HKU1, and HLHK9, which were collected from human clinical samples. The results of the phylogenetic tree suggested that evolutionary divergence between strains of different hosts had occurred in core genes. Different *L. hongkongensis* strains might have their host preference.

For the clinical strains, HKU1 and HLHK9 were closely related with each other, along with PW3643, they were more related with strains derived from frogs, like W71, W105, and W47 in the core IV cluster. HLGZ1 was closely related with W36 and W35 in the core III cluster, indicating that human clinical strains were genetically more related to some strains from frogs. HLGZ1 was collected by a diarrheic human stool sample and reported to be clustered into the same MLST cluster with two frog isolates (Wu et al., 2018). These results were consistent with the hypothesis that some *L. hongkongensis* strains derived from frogs may have pathogenicity potential for humans (Wang et al., 2019).

Virulence Factor Profiles

To further understand pathogenic feathers and measure the disease risk of these *L. hongkongensis* food-associated isolates, we analyzed the virulence factors in each isolate. There were 112 virulence factors detected in total, with an average of about 65 virulence factors among the 49 *L. hongkongensis* strains (**Supplementary Table 4**). They were contributing to invasion, adherence, immune evasion, efflux pump, toxin, motility, stress adaption, and other virulence-related functions of *L. hongkongensis*.

Another phylogenetic tree of 49 *L. hongkongensis* strains was constructed based on sequences of virulence genes (**Figure 3A**). The strains were divided into 4 clusters same as the tree based on core genes, indicating that the evolution of virulence genes was coordinated with the evolution of the core genome in *L. hongkongensis*. Virulence gene variation might play important role in bacterial adaptation to the host environment.

To further understand the virulence-related difference between different sources, a heatmap was drawn to visualize the distribution of the virulence genes (**Figure 3B**). Forty-eight core virulence factors were present in more than 47 strains, and 22 specified virulence factors were only present in one single strain. Strain Y93 and Y56 were distinguished from all other strains for holding 12 and 5 specified virulence factors and with the absence of 14 and 11 core virulence factors. Based on the clustering of samples, clinical strain HLGZ1 and HLHK9 were clustered with W36 and W37, PW3643 and HKU1 were clustered with each other, along with W27, W32, W24, and W33, indicating human clinical strains were more similar with strains isolated from frogs

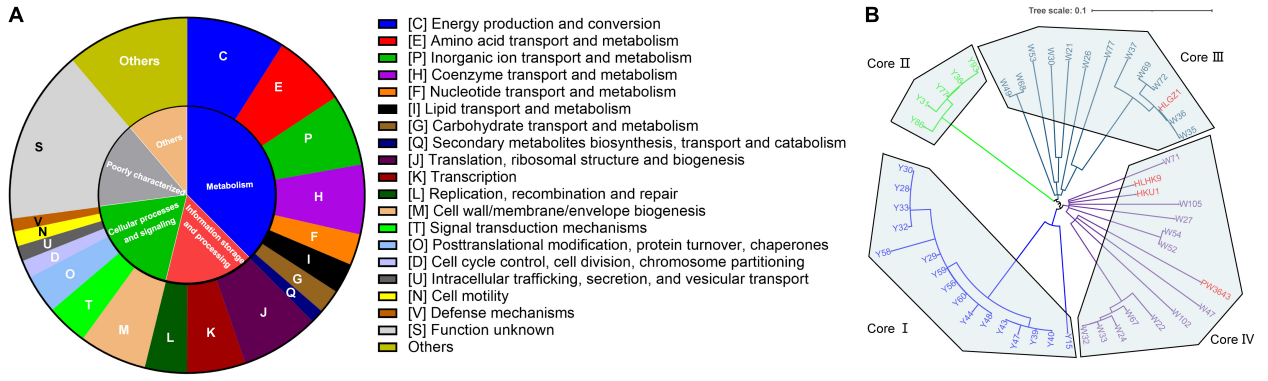


FIGURE 2 | Function annotation and phylogenetic tree of *L. hongkongensis* core genes. **(A)** COG Function categories of core genes. The annotation was conducted with eggNOG based on the COG database. **(B)** Phylogenetic tree of 49 *L. hongkongensis* strains based on amino acid sequences of core genes inferred with PhyloPhlAn 3.0.60. The genes can be grouped into eggNOG orthologous groups, but cannot be grouped into any COG function categories by eggNOG were labeled as "others."

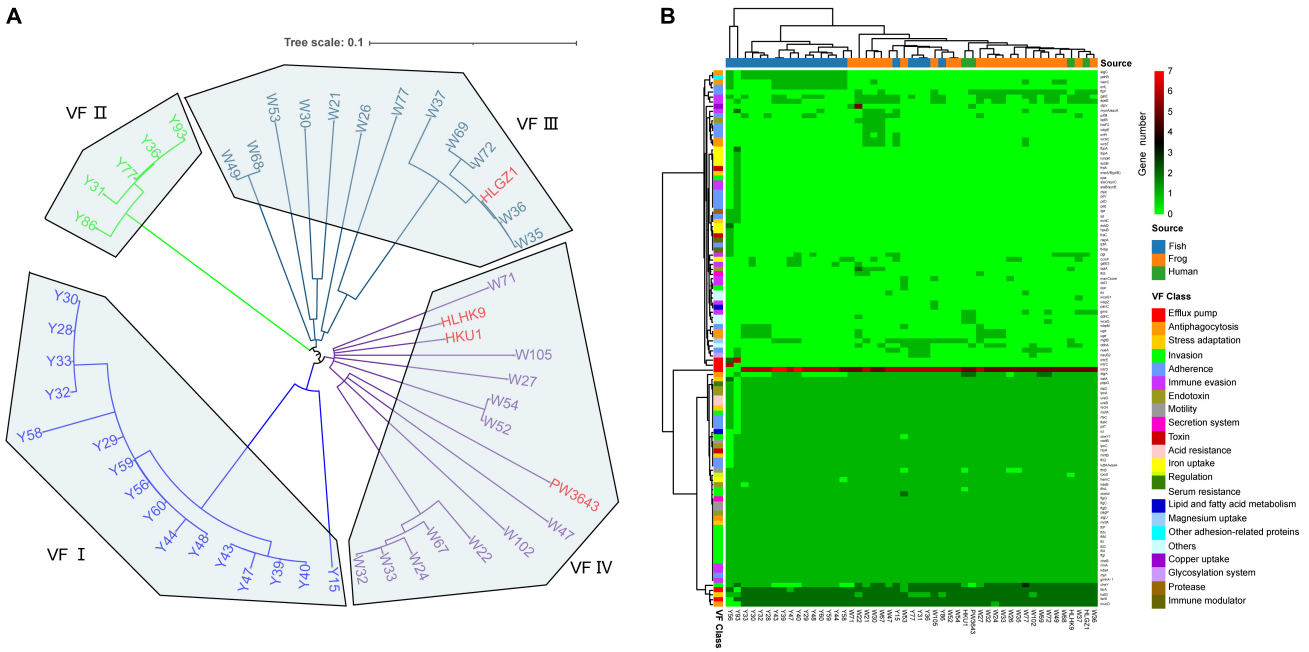


FIGURE 3 | Virulence factor profiles of 49 *L. hongkongensis* genomes. **(A)** Phylogenetic tree based on amino acid sequences of virulence factors using the Maximum Likelihood Method by parsnp. Human clinical strains were presented in strong pink. **(B)** Complex heatmap of predicted virulence genes in 49 genomes. Clustering was performed on row and column distances by ward. D2 method. The gene copy numbers, sample sources, and virulence factor classes were present indicated with different colors.

on virulence factors. This also suggested that strains isolated from frogs may have a higher potential to infect humans, which adds strength to the previous viewpoint (Wang et al., 2019). There were no critical specified VF genes shared by all 4 clinical derived strains, though *gmd* [LPS(Brucella), 3/4 VS 7/49, held by HKU1, HLGZ1, and PW3643], *ddhC* (O-antigen (Yersinia), 2/4 VS 3/49, held by HKU1 and PW3643), and *wcaG* (O-antigen (Yersinia), 2/4 VS 3/49, held by HKU1 and PW3643) were enriched in these human clinical strains, indicating these genes may serve important roles for human infection of *L. hongkongensis*.

Three virulence factors were noticed to be specific of stains from grass carps, *orfL* (adherence), *algC*(Antiphagocytosis), and *pdhB* (other adhesion-related proteins), existing in more than 50% of grass carp derived strains. *AlgC* encodes a phosphoglucomutase, which is essential for biofilm production, and also involved with the formation of protective barrier against host immune defenses and antibiotics and might be important for *L. hongkongensis* infection and survival in grass carps. There was also one gene *wecC* that only existed in grass carp (52.4%) or frog (17.4%) derived strains, and have not been shown in clinical strains, it was known to be involved in the pathway enterobacterial common antigen biosynthesis and in bacterial outer membrane biogenesis (Meier-Dieter et al., 1990). Two genes, *flpF,* and *gmd* were only existed in human (50%) or frog (73.9%) derived strains, and have not shown in grass carp derived strains. One virulence gene, *algA* was found in all strains from frog and human, but only in 47.6% of strains from grass carp, it is known to produce a precursor for alginate polymerization (Shinabarger et al., 1991). The alginate layer provides a protective barrier against host immune defenses and antibiotics and might contribute to survival in the host. The specific virulence-related genes held by strains from different sources might indicate that different strategies have been adopted for multi-host bacteria to adapt to drastic environmental changes.

Carbohydrate-Active enZYme Profiling

Carbohydrate utilization is the core process for organism survival and reflects the adaptability of bacterial strains to environments, and provides hints for understanding the interaction between bacteria and host. CAZyme profiling was analyzed using dbCAN2 to investigate the genomic potential for carbohydrate utilization. The *L. hongkongensis* strains generally had similar types of CAZymes, but there were variations in the absolute numbers of genes within each of certain categories in the CAZy profiles (**Figure 4A** and **Supplementary Table 5**). The mean value of carbohydrate utilization gene number in 49 strains was 78.6. The gene numbers of Auxiliary Activities, Carbohydrate-Binding Modules, Carbohydrate Esterases, and Glycoside Hydrolases were similar between strains from different sources, but strains isolated from frogs and humans carried significantly more genes of Glycosyl Transferases than strains isolated from grass carps (*t*-test $p = 1.49e^{-5}$).

A heatmap was drawn to visualize the presence and absence of Carbohydrate-active enzymes among 49 strains (**Figure 4B**). Eight gene families were present in almost all 49 strains with average gene numbers bigger than 3, indicating their important roles for *L. hongkongensis*. These gene families were AA1 (Levasseur et al., 2013), CBM50 (Boraston et al., 2004), GH23,

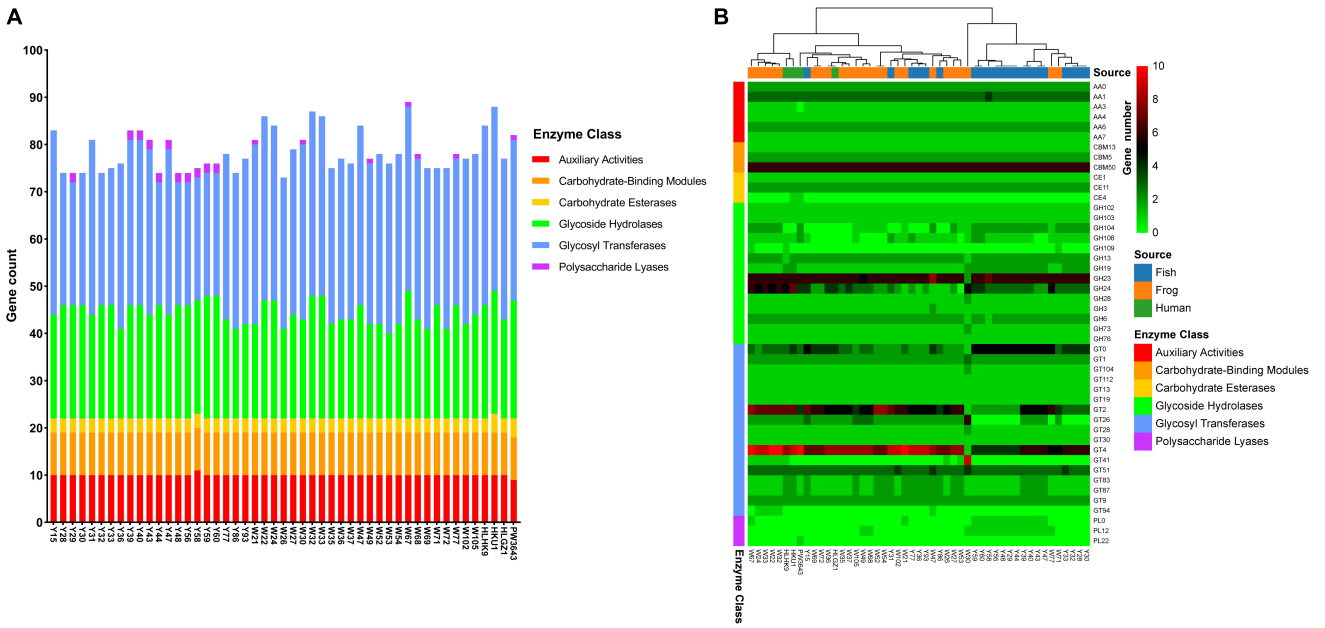


FIGURE 4 | Carbohydrate-Active Enzyme (CAZy) profiles of 49 *L. hongkongensis* genomes. **(A)** Stacked bar chart of Carbohydrate-Active Enzymes distributions. **(B)** Complex heatmap based on gene copy numbers of Carbohydrate-Active Enzymes in each genome. Clustering is only performed on row distance by ward. D2 method. Analysis of the CAZy profiles was annotated using the dbCAN2 resource CAZy family-specific hidden Markov models (HMMs) (Zhang et al., 2018). The sources of strains and the classes of CAZyme modules or domains were represented as colored bars.

GH24 (Henrissat and Davies, 1997), GT0, GT2, GT4, GT51 (Coutinho et al., 2003), which were reported to be multicopper oxidases that use diphenols, enzymes cleaving either chitin or peptidoglycan, lysozymes, cellulose synthase, sucrose synthase, and murein polymerase, and so on.

Human clinical strains seem to have more enzymes of GT4 compared to other strains, the mean value of GT4 enzyme number existed in 4 clinical strains was 8.75, while it was 7.7 in frog-derived strains, and 6.0 in grass carp derived strains. GT4 is the Glycosyltransferase family that contains sucrose synthase, sucrose-phosphate, and other Glycosyltransferases (Campbell et al., 1997), which may be important for *L. hongkongensis* survival inside the human host. PW3643 was the only strain that encoded PL22. PL22s were commonly referred to as oligogalacturonide lyases (OGLs). This enzyme family was found primarily in phytopathogenic or intestinal bacteria where it played a role in the metabolism of pectin (Abbott et al., 2010). PW3643 was also the only strain that do not have an AA3. Enzymes of family AA3 were widespread and catalyzed the oxidation of alcohols or carbohydrates with the concomitant formation of hydrogen peroxide or hydroquinones, the enzymes were most abundant in fungi and only recently found in bacteria (Sützl et al., 2018). CE4 was only found present in PW3643 and HKU1. Known activities of CE 4 family members include acetylxylan esterases, chitin deacetylases, chitooligosaccharide deacetylases, and peptidoglycan deacetylases (Hayhurst et al., 2008). Peptidoglycan was the essential bacterial cell wall polymer, indicating that PW3643 and HKU1 may have specified feathers in their cell wall. Interestingly, we found that clinical strains from Hong Kong, HKU1, and HLHK9 were clustered together with frog-derived strains, strain HLGZ1 was still clustered with W36 and W37 based on CAZyme profiling, consistent with the phylogenetic trees and virulence factors.

We also noticed that only some strains isolated from grass carp uniquely held genes of Polysaccharide Lyases PL0 and PL12, this may be due to the specified polyanionic substrates they need to degrade in their living habitat.

Antimicrobial Resistance Determinants

Antimicrobial resistance is the core focus of clinical treatment and control of pathogenic microorganisms. So, the antibiotic resistance genes were predicted and antimicrobial susceptibility for common drugs were tested. Thirty-two AMR genes or point mutations were identified (**Figure 5A** and **Supplementary Table 6**). Among all detected genes or point mutations, the *L. hongkongensis ampC* which encoded an class C beta-lactamase, and *adeF* which encoded resistance-nodulation-cell division (RND) antibiotic efflux pump had much higher prevalence rates than the other genes or point mutations (both 100%), followed by sulfonamide resistant *sul1* (30.6%), major facilitator superfamily (MFS) antibiotic efflux pump [24.5% *tet(A)*, 10.2% *tet(D)*, 2.0% *tet(C)*, and 2.0% *tet(G)*], aminoglycoside resistant gene *AAC(6′)*(14.3% *AAC(6′)-Ib-cr*, 4.1% *AAC(6′)-Ib9*, 2.0% *AAC(6′)-IIa*, 2.0% *AAC(6′)-Ib′*, and 2.0% *AAC(6′)-Ib7*), small multidrug resistance (SMR) antibiotic efflux pump *qacH* (12.2%), genes that encode proteins for streptomycin nucleotidylylation *ANT(3″)* (10.2% *aadA3*, 8.2% *aadA2*, and 6.1% a*adA16*), genes that encode proteins for streptomycin phosphorylation [4.1% *APH(3″)-Ia*,

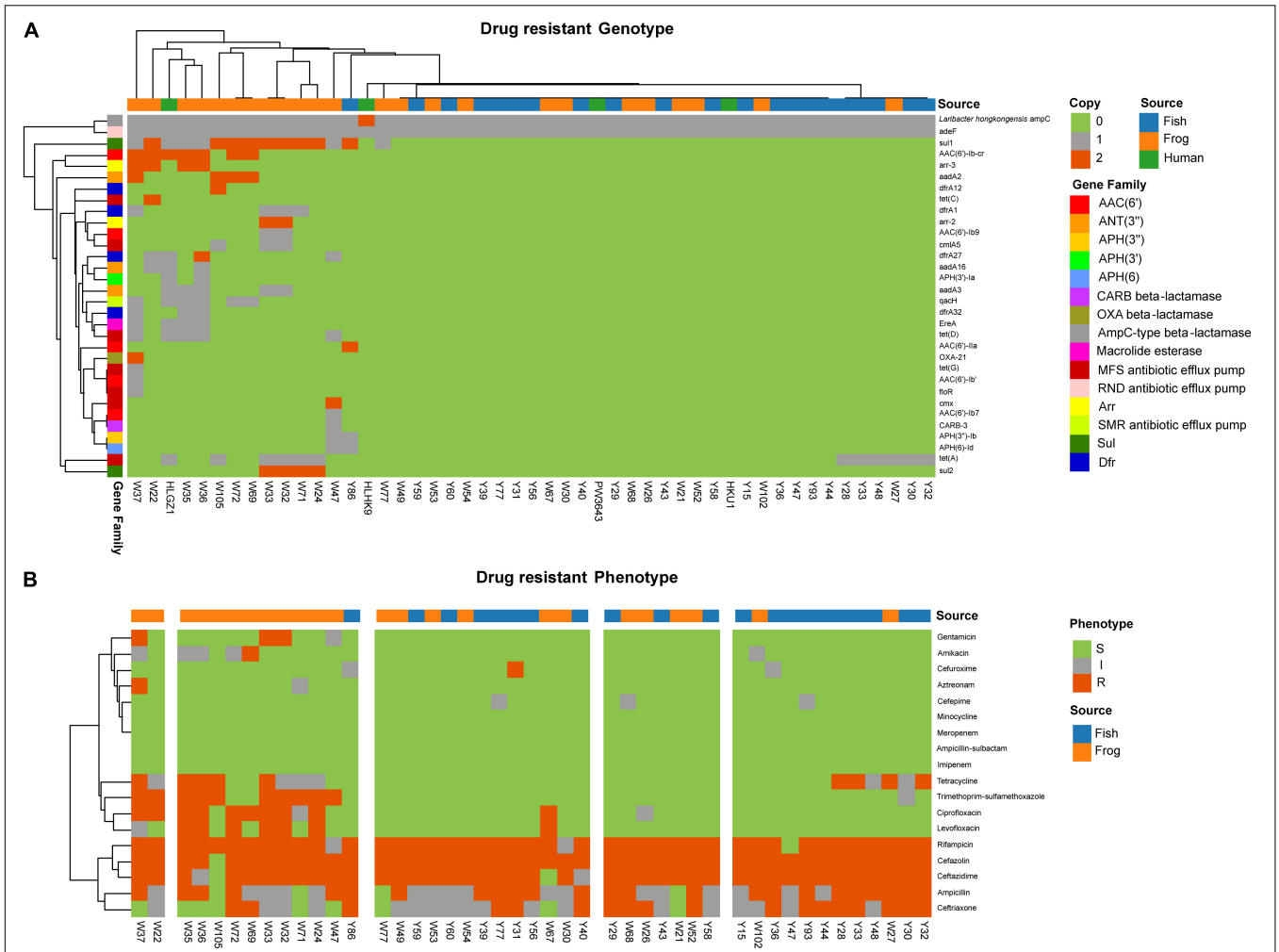


FIGURE 5 | Antibiotic susceptibility and genetic determinants for *L. hongkongensis* strains in this study. **(A)** Genetic determinants of 49 *L. hongkongensis* were displayed in a Complex heatmap. The source of strains, gene copies, and gene families are represented as colored bars. **(B)** Antibiotic susceptibility of 45 aquatic animal isolated strains was displayed, with S indicating susceptible, I indicating intermediate, and R indicating resistance.

4.1% *APH(3″)-Ib*, and 4.1% *APH(6)-Id*], trimethoprim resistant dihydrofolate reductase *dfr* (8.2% *dfrA27*, 8.2% *dfrA1*, and 6.1% *dfrA32*, and 2% *dfrA12*), rifampin ADP-ribosyltransferase (8.2% *arr-3*, 4.1% *arr-2*), macrolide esterase *EreA* (8.2%), genes encoding phenicol antibiotic efflux pump (6.1% *cmlA5*, and 2% *cmx*, and 2% *floR*), and OXA beta-lactamase *OXA-21* (2%), CARB beta-lactamase *CARB-3* (2%).

Susceptibility to 18 clinical commonly used antibiotics was tested for the 45 food-associated strains (**Figure 5B**). *L. hongkongensis* had a general resistance to cefazolin [97.8% Resistance (R) or intermediate (I)], rifampicin (97.8% R or I), ceftazidime (95.6% R or I), ampicillin (91.1% R or I), and ceftriaxone (80.0% R or I), this may be due to the high prevalence rate of *Laribacter hongkongensis ampC*. One-third of strains showed resistance to tetracycline due to *tet(A)* and other *tet* genes. Resistance/intermediate rate of frog-derived strains to fluoroquinolones (Ciprofloxacin, $p = 5.45e^{-5}$, Levofloxacin, $p = 2.83e^{-3}$), aminoglycosides (Amikacin, $p = 1.32e^{-2}$, Gentamicin, $p = 5.15e^{-2}$), Ceftriaxone ($p = 1.20e^{-3}$), and Trimethoprim-sulfamethoxazole ($p = 3.32e^{-3}$) were significantly higher compared to grass carp-derived strains. Among frog-derived strains, 50% of strains were resistant to ciprofloxacin, 33.3% of strains were resistant to levofloxacin, 41.7% of strains were resistant to trimethoprim-sulfamethoxazole, 25% of strains were resistant to amikacin, and 16.7% of strains were resistant to gentamicin. It was also worth mentioning that the resistance/intermediate rate of cefuroxime was higher ($p = 5.71e^{-2}$) in grass carp derived strains (14.3%, 3/21) than in frog-derived strains (0%), which cannot be clearly explained by their AMR gene, this may due to unknown genomic differences in grass carp-derived strains.

The drug resistance of *L. hongkongensis* from frogs and fishes has been reported (Feng et al., 2012; Wang et al., 2019). We believe that the high resistance rates of frog-derived strains to fluoroquinolones and aminoglycosides might

be due to prophylactic use of antibiotics in aquaculture and antibiotic pollution in aquaculture ponds, leading to selection pressure on environmental bacteria to develop resistance and to facilitate horizontal gene transfer (HGT) between bacteria. Sulfametoxydiazine, sulfamethazine, sulfamethoxazole, oxytetracycline, chlortetracycline, doxycycline, ciprofloxacin, norfloxacin, and enrofloxacin had been detected in sediment and water samples in Guangdong province, China (Xiong W. et al., 2015). Fluoroquinolones (Turnipseed et al., 2012) and multidrug-resistant bacteria (Lei et al., 2019) had also been detected from frog tissue. One research had reported that ARG pollution was more serious in bullfrog ponds than polyculture ponds due to the more frequent use of antibiotics in bullfrog rearing operations (Yuan et al., 2019).

The general beta-lactamase resistance of *L. hongkongensis* should be noted and clinical medication strategies adjusted accordingly. The high drug resistance rate of frog-derived strains is becoming a serious problem. We also noticed that drug-resistant frog-derived strains like W35, W36, W37, and so on were closely related with human clinical strains based on phylogenetic trees, virulence factors, and carbohydrate-active enzymes, indicating their high risk of human infection and can bring serious difficulties for clinical treatment of *L. hongkongensis* infection.

CONCLUSION

The genome sequences of *L. hongkongensis* reported here is a valuable resource for further studies investigating the newly discovered foodborne pathogen. The comprehensive analysis of the genetic evolution, virulence factors, carbohydrate-active enzymes can provide a snapshot of the genetic and pathogenic characteristics of *L. hongkongensis*. Prediction of antimicrobial resistance gene and susceptibility analysis of clinical drugs gives high-quality reference information for drug choice, which will benefit the environmental and clinical control of *L. hongkongensis*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Medical Ethics Committee of Zhujiang Hospital, Southern Medical University.

AUTHOR CONTRIBUTIONS

P-BY performed the data analyses and wrote the manuscript. YZ contributed significantly to analysis and manuscript preparation. J-HZ and J-HL performed the experiment. E-ZC, W-TL, L-JW, and Y-XZ helped perform the sample collection, bacteria isolation, data analysis, and constructive discussions. D-QC conceived and designed the study and co-wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The research outlined in this study was supported by the research grants from the Natural Science Foundation of China (No. 81974318), the Natural Science Foundation of Guangdong Province (No. 2018A030313279), and the Guangdong Province Science and Technology Innovation Strategy Special Fund (No. 2019B020209001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.862776/full#supplementary-material

Supplementary Table 1 | The general characteristics of *L. hongkongensis genomes*.

Supplementary Table 2 | The gene presence and absence matrix and annotation of all identified genes.

Supplementary Table 3 | The functions of core genes were characterized by matching the sequences with the COG database.

Supplementary Table 4 | Virulence factors detected in the 49 *L. hongkongensis* strains.

Supplementary Table 5 | CAZyme profiling of the 49 *L. hongkongensis* strains.

Supplementary Table 6 | Antibiotic resistance genes predicted and antimicrobial susceptibility for drugs of *L. hongkongensis* strains.

REFERENCES

Abbott, D. W., Gilbert, H. J., and Boraston, A. B. (2010). The active site of Oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate beta-elimination. *J. Biol. Chem.* 285, 39029–39038. doi: 10.1074/jbc.M110.153981

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* 11:2500. doi: 10.1038/s41467-020-16366-7

Beilfuss, H. A., Quig, D., Block, M. A., and Schreckenberger, P. C. (2015). Definitive identification of *Laribacter hongkongensis* acquired in the United States. *J. Clin. Microbiol.* 53, 2385–2388. doi: 10.1128/jcm.00539-15

Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781. doi: 10.1042/BJ20040892

Buchfink, B., Reuter, K., and Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368. doi: 10.1038/s41592-021-01101-x

Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* 326(Pt. 3), 929–939. doi: 10.1042/bj3260929u

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* 328, 307–317. doi: 10.1016/s0022-2836(03)00307-3

Engsbro, A. L., Nielsen, K. L., Hornum, M., and Andersen, L. P. (2018). *Laribacter hongkongensis*: clinical presentation, epidemiology and treatment. A review of the literature and report of the first case in Denmark. *Infect. Dis. (Lond. Engl.)* 50, 417–422. doi: 10.1080/23744235.2017.1419373

Feng, J. L., Hu, J., Lin, J. Y., Liu, S., Chowdhury, N., Zhang, O., et al. (2012). The prevalence, antimicrobial resistance and PFGE profiles of *Laribacter hongkongensis* in retail freshwater fish and edible frogs of southern China. *Food Microbiol.* 32, 118–123. doi: 10.1016/j.fm.2012.04.018

Greay, T. L., Gofton, A. W., Zahedi, A., Paparini, A., Linge, K. L., Joll, C. A., et al. (2019). Evaluation of 16S next-generation sequencing of hypervariable region 4 in wastewater samples: an unsuitable approach for bacterial enteric pathogen identification. *Sci. Total Environ.* 670, 1111–1124. doi: 10.1016/j.scitotenv.2019.03.278

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hayhurst, E. J., Kailas, L., Hobbs, J. K., and Foster, S. J. (2008). Cell wall peptidoglycan architecture in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14603–14608. doi: 10.1073/pnas.0804138105

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. doi: 10.1016/s0959-440x(97)80072-3

Hernandez, D., Tewhey, R., Veyrieras, J. B., Farinelli, L., Østerås, M., François, P., et al. (2014). De novo finished 2.8 Mbp *Staphylococcus aureus* genome assembly from 100 bp short and long range paired-end reads. *Bioinformatics (Oxford, England)* 30, 40–49. doi: 10.1093/bioinformatics/btt590

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hung, D. L. L., Teng, J. L. L., Fong, J. Y. H., Wang, Q., Chen, Z., Fung, A. M. Y., et al. (2020). Severe underlying liver diseases and high mortality associated with *Laribacter hongkongensis* bacteremia. *Diagn. Microbiol. Infect. Dis.* 96:114948. doi: 10.1016/j.diagmicrobio.2019.114948

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, D. S., Wi, Y. M., Choi, J. Y., Peck, K. R., Song, J. H., and Ko, K. S. (2011). Bacteremia caused by *Laribacter hongkongensis* misidentified as *Acinetobacter lwoffii*: report of the first case in Korea. *J. Korean Med. Sci.* 26, 679–681. doi: 10.3346/jkms.2011.26.5.679

Kong, H. K., Law, H. W., Liu, X., Law, C. O., Pan, Q., Gao, L., et al. (2017). Transcriptomic analysis of *Laribacter hongkongensis* reveals adaptive response coupled with temperature. *PLoS One* 12:e0169998. doi: 10.1371/journal.pone.0169998

Kong, Q., Sun, J., Shen, L., Cha, J., Xu, H., Jin, H., et al. (2016). Investigation on the effect of the ecological parameters on the prevalence of *Laribacter hongkongensis* in freshwater fish and in human. *Indian J. Med. Microbiol.* 34, 110–111. doi: 10.4103/0255-0857.167670

Lau, S. K., Lee, L. C., Fan, R. Y., Teng, J. L., Tse, C. W., Woo, P. C., et al. (2009). Isolation of *Laribacter hongkongensis*, a novel bacterium associated with gastroenteritis, from Chinese tiger frog. *Int. J. Food Microbiol.* 129, 78–82. doi: 10.1016/j.ijfoodmicro.2008.10.021

Lau, S. K., Woo, P. C., Fan, R. Y., Lee, R. C., Teng, J. L., and Yuen, K. Y. (2007a). Seasonal and tissue distribution of *Laribacter hongkongensis*, a novel bacterium associated with gastroenteritis, in retail freshwater fish in Hong Kong. *Int. J. Food Microbiol.* 113, 62–66. doi: 10.1016/j.ijfoodmicro.2006.07.017

Lau, S. K., Woo, P. C., Fan, R. Y., Ma, S. S., Hui, W. T., Au, S. Y., et al. (2007b). Isolation of *Laribacter hongkongensis*, a novel bacterium associated with gastroenteritis, from drinking water reservoirs in Hong Kong. *J. Appl. Microbiol.* 103, 507–515. doi: 10.1111/j.1365-2672.2006.03263.x

Lei, X. P., Yi, G., Wang, K. Y., OuYang, P., Chen, F., Huang, X. L., et al. (2019). *Elizabethkingia miricola* infection in Chinese spiny frog (*Quasipaa spinosa*). *Transbound. Emerg. Dis.* 66, 1049–1053. doi: 10.1111/tbed.13101

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6:41. doi: 10.1186/1754-6834-6-41

Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Meier-Dieter, U., Starman, R., Barr, K., Mayer, H., and Rick, P. D. (1990). Biosynthesis of enterobacterial common antigen in *Escherichia coli*. Biochemical characterization of Tn10 insertion mutants defective in enterobacterial common antigen synthesis. *J. Biol. Chem.* 265, 13490–13497. doi: 10.1016/s0021-9258(18)77373-0

Ni, X., Sun, J., Kong, Q., Kong, F., Brown, M., Shen, L., et al. (2011). Isolation of *Laribacter hongkongensis* from little egrets (*Egretta garzetta*) in Hangzhou, China. *Lett. Appl. Microbiol.* 52, 465–467. doi: 10.1111/j.1472-765X.2011.03024.x

Nourdin-Galindo, G., Sánchez, P., Molina, C. F., Espinoza-Rojas, D. A., Oliver, C., Ruiz, P., et al. (2017). Comparative pan-genome analysis of *Piscirickettsia salmonis* reveals genomic divergences within genogroups. *Front. Cell. Infect. Microbiol.* 7:459. doi: 10.3389/fcimb.2017.00459

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics (Oxford, England)* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Raja, K. M., and Ghosh, A. R. (2014). Molecular insight of putative pathogenicity markers with ESBL genes and lipopolysaccharide in *Laribacter hongkongensis*. *Appl. Biochem. Biotechnol.* 174, 1935–1944. doi: 10.1007/s12010-014-1163-0

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shinabarger, D., Berry, A., May, T. B., Rothmel, R., Fialho, A., and Chakrabarty, A. M. (1991). Purification and characterization of phosphomannose isomerase-guanosine diphospho-D-mannose pyrophosphorylase. A bifunctional enzyme in the alginate biosynthetic pathway of *Pseudomonas aeruginosa*. *J. Biol. Chem.* 266, 2080–2088. doi: 10.1016/s0021-9258(18)52212-2

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sützl, L., Laurent, C., Abrera, A. T., Schütz, G., Ludwig, R., and Haltrich, D. (2018). Multiplicity of enzymatic functions in the CAZy AA3 family. *Appl. Microbiol. Biotechnol.* 102, 2477–2492. doi: 10.1007/s00253-018-8784-0

Teng, J. L. L., Luo, R., Tang, B. S. F., Fong, J. Y. H., Wang, L., Jia, L., et al. (2021). High prevalence and mechanism associated with extended spectrum beta-lactamase-positive phenotype in *Laribacter hongkongensis*. *Front. Microbiol.* 12:618894. doi: 10.3389/fmicb.2021.618894

Teng, J. L., Woo, P. C., Ma, S. S., Sit, T. H., Ng, L. T., Hui, W. T., et al. (2005). Ecoepidemiology of *Laribacter hongkongensis*, a novel bacterium associated with gastroenteritis. *J. Clin. Microbiol.* 43, 919–922. doi: 10.1128/JCM.43.2.919-922.2005

Tettelin, H., Masignani, 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

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/s13059-014-0524-x

Tse, C. W., Curreem, S. O., Cheung, I., Tang, B. S., Leung, K. W., Lau, S. K., et al. (2014). A novel MLST sequence type discovered in the first fatal case of *Laribacter hongkongensis* bacteremia clusters with the sequence types of other human isolates. *Emerg. Microbes Infect.* 3:e41. doi: 10.1038/emi.2014.39

Turnipseed, S. B., Clark, S. B., Storey, J. M., and Carr, J. R. (2012). Analysis of veterinary drug residues in frog legs and other aquacultured species using liquid chromatography quadrupole time-of-flight mass spectrometry. *J. Agric. Food Chem.* 60, 4430–4439. doi: 10.1021/jf2049905

Wang, L., Fu, L., Liu, Z., Guo, H., Wang, L., Feng, M., et al. (2019). Comparative analysis of antimicrobial resistance, integrons, and virulence genes among extended-spectrum β-lactamase-positive *Laribacter hongkongensis* from edible frogs and freshwater fish. *Microb. Drug Resist. (Larchmont, NY)* 25, 855–864. doi: 10.1089/mdr.2018.0366

Wang, Z., Zhu, J., Liu, Z., Liu, Y., Zheng, N., Feng, M., et al. (2017). Multi-locus sequence typing of *Laribacter hongkongensis* isolates from freshwater animals, environment and diarrhea patients in southern China. *Int. J. Food Microbiol.* 245, 98–104. doi: 10.1016/j.ijfoodmicro.2017.01.016

Woo, P. C., Lau, S. K., Teng, J. L., and Yuen, K. Y. (2005). Current status and future directions for *Laribacter hongkongensis*, a novel bacterium associated with gastroenteritis and traveller's Diarrhoea. *Curr. Opin. Infect. Dis.* 18, 413–419. doi: 10.1097/01.qco.0000180162.76648.c9

Woo, P. C., Lau, S. K., Teng, J. L., Que, T. L., Yung, R. W., Luk, W. K., et al. (2004). Association of *Laribacter hongkongensis* in community-acquired gastroenteritis with travel and eating fish: a multicentre case-control study. *Lancet (London, England)* 363, 1941–1947. doi: 10.1016/S0140-6736(04)16407-6

Wu, H. K., Chen, J. H., Yang, L., Li, A. R., Su, D. H., Lin, Y. P., et al. (2018). Emergence and genomic analysis of MDR *Laribacter hongkongensis* strain HLGZ1 from Guangzhou, China. *J. Antimicrob. Chemother.* 73, 643–647. doi: 10.1093/jac/dkx470

Xie, J., He, J. B., Shi, J. W., Xiao, Q., Li, L., and Woo, P. C. (2014). An adult zebrafish model for *Laribacter hongkongensis* infection: Koch's postulates fulfilled. *Emerg. Microbes Infect.* 3:e73. doi: 10.1038/emi.2014.73

Xiong, L., Chan, E., Teng, J. L. L., Liu, S., Lau, S. K. P., and Woo, P. C. Y. (2019). Malate-dependent carbon utilization enhances central metabolism and contributes to biological fitness of *Laribacter hongkongensis* via CRP regulation. *Front. Microbiol.* 10:1991. doi: 10.3389/fmicb.2019.01991

Xiong, L., Teng, J. L., Watt, R. M., Liu, C., Lau, S. K., and Woo, P. C. (2015). Molecular characterization of arginine deiminase pathway in *Laribacter hongkongensis* and unique regulation of arginine catabolism and anabolism by multiple environmental stresses. *Environ. Microbiol.* 17, 4469–4483. doi: 10.1111/1462-2920.12897

Xiong, L., Yang, Y., Ye, Y. N., Teng, J. L., Chan, E., Watt, R. M., et al. (2017). *Laribacter hongkongensis* anaerobic adaptation mediated by arginine metabolism is controlled by the cooperation of FNR and ArgR. *Environ. Microbiol.* 19, 1266–1280. doi: 10.1111/1462-2920.13657

Xiong, W., Sun, Y., Zhang, T., Ding, X., Li, Y., Wang, M., et al. (2015). Antibiotics, antibiotic resistance genes, and bacterial community composition in fresh water aquaculture environment in China. *Microb. Ecol.* 70, 425–432. doi: 10.1007/s00248-015-0583-x

Yuan, K., Wang, X., Chen, X., Zhao, Z., Fang, L., Chen, B., et al. (2019). Occurrence of antibiotic resistance genes in extracellular and intracellular DNA from sediments collected from two types of aquaculture farms. *Chemosphere* 234, 520–527. doi: 10.1016/j.chemosphere.2019.06.085

Yuen, K. Y., Woo, P. C., Teng, J. L., Leung, K. W., Wong, M. K., and Lau, S. K. (2001). *Laribacter hongkongensis* gen. nov., sp. nov., a novel gram-negative bacterium isolated from a cirrhotic patient with bacteremia and empyema. *J. Clin. Microbiol.* 39, 4227–4232. doi: 10.1128/JCM.39.12.4227-4232.2001

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fmicb.2022.858983

The Minimal Translation Machinery: What We Can Learn From Naturally and Experimentally Reduced Genomes

*María José Garzón[1], Mariana Reyes-Prieto[2,3] and Rosario Gil[1,2]**

[1] Departament de Genètica, Universitat de València, Burjassot, Spain, [2] Institute for Integrative Systems Biology, Universitat de València–Consejo Superior de Investigaciones Científicas, Paterna, Spain, [3] Sequencing and Bioinformatics Service, Foundation for the Promotion of Sanitary and Biomedical Research of the Valencian Community, Valencia, Spain

The current theoretical proposals of minimal genomes have not attempted to outline the essential machinery for proper translation in cells. Here, we present a proposal of a minimal translation machinery based on (1) a comparative analysis of bacterial genomes of insects' endosymbionts using a machine learning classification algorithm, (2) the empiric genomic information obtained from *Mycoplasma mycoides* JCVI-syn3.0 the first minimal bacterial genome obtained by design and synthesis, and (3) a detailed functional analysis of the candidate genes based on essentiality according to the DEG database (*Escherichia coli* and *Bacillus subtilis*) and the literature. This proposed minimal translational machinery is composed by 142 genes which must be present in any synthetic prokaryotic cell designed for biotechnological purposes, 76.8% of which are shared with JCVI-syn3.0. Eight additional genes were manually included in the proposal for a proper and efficient translation.

Keywords: translation machinery, minimal genome, endosymbionts, JCVI-sync3.0, cosymbionts

OPEN ACCESS

Edited by:
Feng Gao,
Tianjin University, China

Reviewed by:
Daniel R. Zeigler,
The Ohio State University, United States
Johannes Kabisch,
Norwegian University of Science and Technology, Norway

***Correspondence:**
Rosario Gil
rosario.gil@uv.es

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

Received: *20 January 2022*
Accepted: *17 March 2022*
Published: *11 April 2022*

Citation:
Garzón MJ, Reyes-Prieto M and Gil R (2022) The Minimal Translation Machinery: What We Can Learn From Naturally and Experimentally Reduced Genomes. Front. Microbiol. 13:858983. doi: 10.3389/fmicb.2022.858983

INTRODUCTION

The minimal genome was originally defined as the set of genes necessary and sufficient for life under low restrictive conditions (Mushegian, 1999). Therefore, a minimal genome must guarantee the three functional pillars of a living cell (Gil, 2014). First, a simplified DNA replication and repair system, as well as transcription and translation systems, to ensure the maintenance and the proper use of its genetic information; second, a self-sufficient metabolism that meets basic energy and structural requirements; last, an envelope that shelters all the cellular machinery and allows interaction with the environment, as well as the generation of descendants.

Many authors have striven to define the minimal genome to understand the basic principles of life and to apply it for biotechnological purposes (Gil, 2014; Hutchison et al., 2016; Ziegler and Takors, 2019). One of the most used strategies to define the minimal genome is based on the study of organisms with naturally reduced genomes due to their living style in association with a eukaryotic host, no matter if the association is mutualistic or parasitic (Moya et al., 2008). On account of this host-dependence characteristic and the niches in which they thrive, it has always been a challenge to apply experimental techniques to investigate their physiological processes in real-time. Nevertheless, thanks to the development of high-throughput sequencing technologies,

many genomes of mutualistic insect endosymbionts have been sequenced and are available for in depth studies. These bacteria have been recognized as a key factor in the evolutionary success of this group of animals (Moya et al., 2008), as they provide their hosts with tools to adapt to new environments, in many cases related to nutrient provision. During the process of symbiotic integration, the endosymbiont genomes undergo what has been called the "genome-reduction syndrome" (Gil et al., 2010). The genome size of obligate (primary) endosymbionts (OS) can vary, depending on the age of the association and the degree of symbiotic integration achieved, leading to small genomes (circa 600 kb in many cases) or even to tiny genomes (Moran and Bennett, 2014), also known as "symbionelles" (Reyes-Prieto et al., 2014), as it is the case of the 106-kb genome of "*Candidatus* Hodgkinia cicadicola" str. TETCHI4, the smallest sequenced bacterial genome to date (Łukasik et al., 2018). The strong genome shrinkage undergone by these tiny genomes, has gone beyond the limit of what has been defined as a minimal genome (Gil, 2014), which contemplates about 187–205 protein-coding and 35–38 RNA genes. Although this proposal includes universally retained genes, and some persistent ones [i.e., non-ubiquitous genes conserved in most genomes, therefore, non-essential but needed for robust long-term survival; Acevedo-Rocha et al. (2013)], it is still a theoretical proposal that has never been proven to be enough to maintain a living cell.

The translational machinery is essential for the maintenance and continuity of the cell and is, by far, the most complex part of modern cells (**Figure 1**). It is made of numerous macromolecules, including proteins and RNAs (mRNAs, tRNAs, rRNAs, and other small RNAs), all of which are encoded in an organism's genome. Most genes involved in translation are considered essential or quasi-essential for cell survival. Remarkably, there are many examples of endosymbiotic bacteria that lack important translational genes. In such cases, it has been hypothesized that it must be the host, or a co-obligate endosymbiont where appropriate, which provides the informational precursors to the endosymbiont which holds the deficiency (Sloan and Moran, 2012).

In 2010, the 1,079-kb genome of *M. mycoides* JCVI-syn1.0 was chemically synthesized and its cell growth when transplanted into the cytoplasm of *Mycoplasma capricolum* was proven (Gibson et al., 2010). The first semisynthetic organism based on modern living cells, was created. Afterwards, using JCVI-syn1.0 as a starting point, and by removing non-essential genomic regions through a cyclic design-build-test (DBT) strategy, Hutchison and co-workers managed to obtain *M. mycoides* JCVI-syn3.0 (Hutchison et al., 2016). This semisynthetic organism is viable in axenic culture and its streamlined genome contains 438 protein-coding and 35 RNA genes. It derives from a Mollicutes, which have evolved from ancestral Gram-positive bacteria (Parks et al., 2018) for which the translation machinery has been extensively studied in the last decade (Grosjean et al., 2014). In their work, Grosjean and collaborators identified translation-related protein-coding genes shared by 39 selected Mollicutes' genomes and compared them with those of *Escherichia coli* and *Bacillus subtilis*, as Gram-negative and Gram-positive bacterial models, respectively. A set of 260 protein-coding genes involved in translational functions were selected for the study and classified in the following functional categories: ribosomal proteins, tRNA aminoacylation, rRNA modifications, tRNA modifications, ribosome assembly, translation, and RNA processing. They found that the categories of aminoacyl-tRNA synthetases, ribosomal proteins and translation factors contained the most preserved genes, while some enzymes involved in specific modifications of tRNAs, 16S rRNA, and 23S rRNA, fundamental for decoding and peptidyl transfer, were also essential.

The updated proposal of the minimal genome by Gil (2014) included a revised version of the genes involved in translation. It is worth mentioning that, among the genes included in the previously defined core of the minimal genome (Gil et al., 2004), four of the poorly characterized genes due to the lack of information at that time, were later identified as part of the translation machinery (i.e., *rsm*H, *rsm*I, *til*S, and *ybe*Y). Twelve additional persistent genes were added to the new version, many of which encode ribosomal proteins that, due to their small size, might have been missed during genome annotation or lost in extremely reduced genomes (Nikolaeva et al., 2021).

For this work, in an attempt to get closer to the universal core of the minimal translation machinery, we selected most of the complete reduced genomes of insect endosymbionts available in the SymGenDB database by 2020 (Reyes-Prieto et al., 2020), to search through an unsupervised machine learning technique (hierarchical clustering) for those essential and persistent genes involved in translation. Then, in order to validate our *in silico* minimal translation machinery proposal, we compared it with that of *M. mycoides* JCVI-syn3.0. Finally, we compared the obtained translation machinery with the one defined by Grosjean et al. (2014) for Mollicutes. Our final proposal is composed of 142 protein-coding genes and defines the protein components of the minimal translation machinery that must be present in a hypothetical viable prokaryotic cell, which can be useful for defining a biological chassis to which desired functions can be added for biotechnological purposes.

MATERIALS AND METHODS

Bacterial Genomes Used in This Study

The endosymbiont genomes to be included in the analyses and their accession IDs were mostly retrieved from SymGenDB (Reyes-Prieto et al., 2020). We added the following genomes that were not available in SymGenDB (2020) due to their posterior discovery or annotation, or because they are not listed in the KEGG database: "*Candidatus* Serratia symbiotica" SeCistrobi, "*Candidatus* Tremblaya phenacola" PPER, "*Candidatus* Tremblaya princeps" TPPLON, *Cardinium* cSfur, *Cardinium hertigii* cBtQ1, *Neisseria meningitidis* MC58, *Serratia symbiotica* SCt-VLC, "*Candidatus* Tremblaya princeps" TPMHIR1, "*Candidatus* Tremblaya princeps" TPPMAR1, "*Candidatus* Tremblaya princeps" TPFVIR, "*Candidatus* Tremblaya princeps" TPTPER1, "*Candidatus* Sulcia muelleri" TETUND, "*Candidatus* Hodgkinia cicadicola" TETUND2, *Buchnera aphidicola* BCc and "*Candidatus* Sodalis sp." SoCistrobi. Their genomic data were retrieved automatically on August 2020 using the efetch

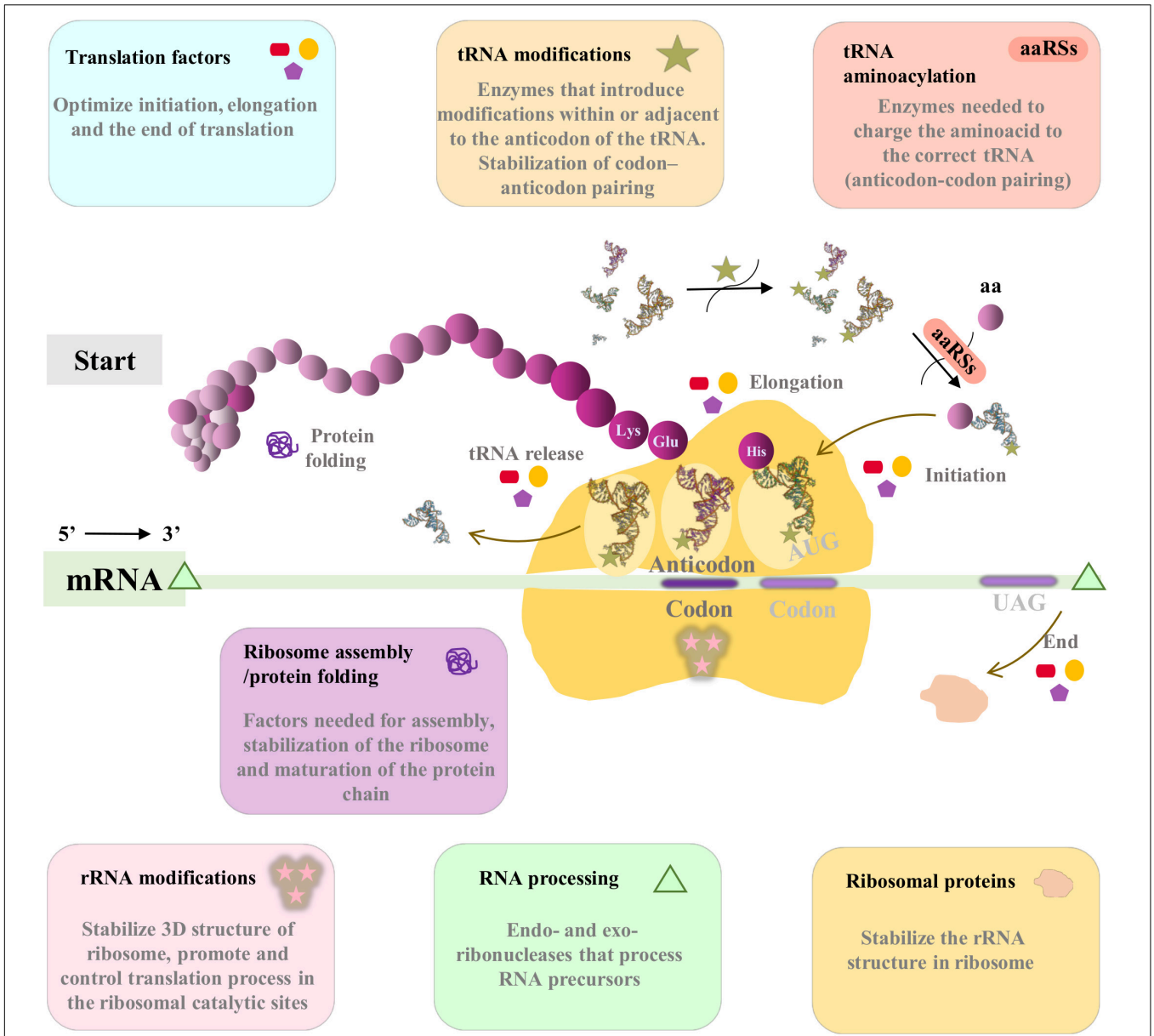


FIGURE 1 | Simplified schematic view of translation in bacteria. In this work all the proteins included in translation machinery are classified in seven functional categories: ribosomal proteins (orange), tRNA aminoacylation (coral), rRNA modifications (pink), tRNA modifications (wheat), ribosome assembly/protein folding (orchid), Translation factors (blue), and RNA processing (green). The most important functions are summarized in the respective panels of each category. The tRNA structure corresponds to tRNA(Ile) from PDB database (1FFY; Silvian et al., 1999), and has been modified with PyMOL software (Schrödinger and DeLano, 2020) for this figure.

command from GeneBank,[1] except for "*Candidatus* Tremblaya phenacola" PPER, *Serratia symbiotica* SCt-VLC and *Cardinium hertigii* cBtQ1, which were manually downloaded because only shotgun assemblies are available. Additionally, in order to generate a complete universe of translation-involved genes, we included in our analysis 10 bacterial genomes with no reduction, two of them, *E. coli* and *B. subtilis*, are common bacterial models for Gram-negative and Gram-positive bacteria, respectively. The other eight were selected because they are taxonomically diverse and can be grown in the laboratory in axenic culture. The Prokka software tool (Seemann, 2014) was used to re-annotate all genomes for homogeneous results. Finally, we also included in our comparisons the genome of *M. mycoides* JCVI-syn3.0. This genome annotation was retrieved from Hutchison et al. (2016). We only took into consideration genes classified as ribosome biogenesis, RNA metabolism, protein folding, translation, RNA, rRNA modification, tRNA modification, and regulation. The sequences of ORFs classified

[1]https://www.ncbi.nlm.nih.gov/genome/

as "unclear category" were used to perform a BLASTP against the non-redundant protein sequences database at the NCBI web (The Blast Sequence Analysis Tool, 2022) to look for putative functions of the hypothetical conserved proteins they might encode. All 110 bacterial genomes under study have been compiled in a dataset called "cosym" with 92 entries as a result of considering coprimary insect endosymbionts (symbiotic consortia) as single entities (**Supplementary Material 1**), and a list of the genera included in this study is listed in **Table 1**.

Identification of the Translational Gene Sets and Gene Orthology Analysis

Genes were classified into functional categories based on the previous work by Grosjean et al. (2014). We defined the set of translational protein-coding genes (*translational gene set*) of the two model species considered through several steps. First, we retrieved the genes from *E. coli* and *B. subtilis* that have been included in the work by Grosjean et al. (2014) as encoding essential components of the translation machinery. Then, we searched for selected GeneOntology (GO) terms on UniProt (UniProt Consortium, 2015) and EcoCyc (Keseler et al., 2013), in order to update the list (**Supplementary Material 2**). The GO terms included were: 0000154 (rRNA modification), 0001510 (RNA methylation), 0001680 (Addition of CCA 3'-end of tRNA), 0005840 (Ribosome), 0006364 (rRNA processing), 0006396 (RNA processing), 0006400 (tRNA modification), 0006412 (Translation), 0006417 (Translational regulation), 0006457 (Protein folding), 0008033 (tRNA processing), 0009451 (RNA modification), 0042255 (Ribosomal assembly), and 0042254 (Ribosomal biogenesis). In this step, a manual curation of gene names was mandatory to eliminate duplicated candidates (i.e., the same ortholog with different annotated names), to use the UniProt accepted nomenclature for genes with a double translational function, and to remove genes not strictly related to translation.

Orthologous genes (paralogs included) in all genomes under study were identified using the Roary software (Page et al., 2015) with default parameters. The absence (0 count) or presence (1 count) of each one of these orthologs in each genome was counted, creating several matrices for further analyses, one per each functional translation category (**Supplementary Material 3**).

TABLE 1 | Genera of the symbionts and free-living bacteria whose genomes have been used in this study.

Symbionts	
Arsenophonus (1)	*Riesia* (2)
Baumannia (2)	*Serratia* (3)
Blattabacterium (8)	*Sodalis* (3)
Blochmannia (6)	*Sulcia* (8)
Buchnera (19)	*Tachikawaea* (1)
Carsonella (7)	*Tremblaya* (8)
Doolittlea (1)	*Uzinura* (1)
Evansia (1)	*Walczuchella* (1)
Gullanella (1)	*Zinderia* (1)
Hoaglandella (1)	*Cardinium* (3)
Hodgkinia (3)	*Legionella* (1)
Mikella (1)	*Neisseria* (1)
Moranella (1)	*Rickettsia* (1)
Nasuia (1)	*Wigglesworthia* (2)
Pantoea (1)	*Wolbachia* (6)
Profftella (1)	Non-genera, Secondary endosymbiont of (3)
Free-living	
Achromobacter (1)	*Mesorhizobium* (1)
Bacillus (1)	*Pseudomonas* (1)
Caulobacter (1)	*Serratia* (1)
Escherichia (1)	*Sphingobacterium* (1)
Flavobacterium (2)	

In brackets, the number of genomes within the same genera.

Classification of Genes and Statistical Analysis

Hierarchical cluster analysis (HCA), an unsupervised machine learning approach for grouping datasets into clusters, was used for the classification of the dataset genes. HCA was performed in R with gplots:heatmap.2 and stats:hclust (complete method) functions. The input data were the presence/absence matrices obtained in the gene orthology analysis. To extract the gene names of the dendrogram clustering, the cutree function of the stats R-package was used. The treatment of the data, the construction of figures and the statistical analyses performed for this work were carried out with *ad doc* scripting in RStudio 4.0.3 using the stats, ggplot2 (Wickham, 2016) and ggthemes R packages. The full script is available through https://github.com/majogarzon/MinTransMach.git.

RESULTS

Selection of the Bacterial Genomes for the Study and Characterization of Their Pangenome

Defining the basic living functions to generate a simplified bacterium that might be modulated under laboratory conditions with desirable and predictable outcomes for biotechnological purposes is a great challenge. The naturally reduced genomes of insect endosymbionts have historically been studied to approach the minimal genome concept, providing valuable information about those functional modules that are necessary and sufficient for life. Previous comparative studies (Gil et al., 2004) concluded that the minimal genome is substantially enriched in genes involved in genetic-information processing, mainly coding for the elements of the translational apparatus, the most complex machinery in a living cell. Yet, it has been possible to define simplified but still functional translation machineries after a reductive evolution in Mollicutes and insect endosymbionts (Grosjean et al., 2014; Gil and Peretó, 2015).

In this work, to further explore and validate the minimal translation machinery, our search began with the selection of organisms with naturally reduced genomes to compare them

with known free-living bacterial models which must possess efficient, complete, and more complex translational apparatus. As a starting point, taking advantage of the availability of extensive genomic information and the bioinformatic tools developed in recent years, we retrieved all the genomes annotated as insect endosymbionts from SymGenDB, a database that lodges genomic information of organisms involved in symbiotic relationships (Reyes-Prieto et al., 2020), and updated the information by manually including some additional endosymbiont genomes, as described in Materials and Methods. All these genomes were classified depending on their symbiotic relationship as primary or obligate symbionts (OS) when they are necessary for the survival and reproduction of the host, and as secondary symbionts (SS) when they maintain a facultative symbiotic relationship in terms of survival. Furthermore, we included several non-symbiotic organisms, designated as free-living (FL), to have a complete representation of the universe of translational genes in our data set. All the 110 organisms used in our analyses are listed along with taxonomic and genomic information in **Supplementary Material 1**, and a summary of their genera is listed in **Table 1**.

Many insects live in obligate association with more than one endosymbiotic bacterium. It has been observed that the presence of two (or more) co-primary endosymbionts allows a greater reduction of the bacterial genomes, far below the definition of a minimal genome (Sloan and Moran, 2012). Probably this means that they can complement each other by exchanging some gene products to perform essential functions (Reyes-Prieto et al., 2014), including informational ones (i.e., DNA replication,

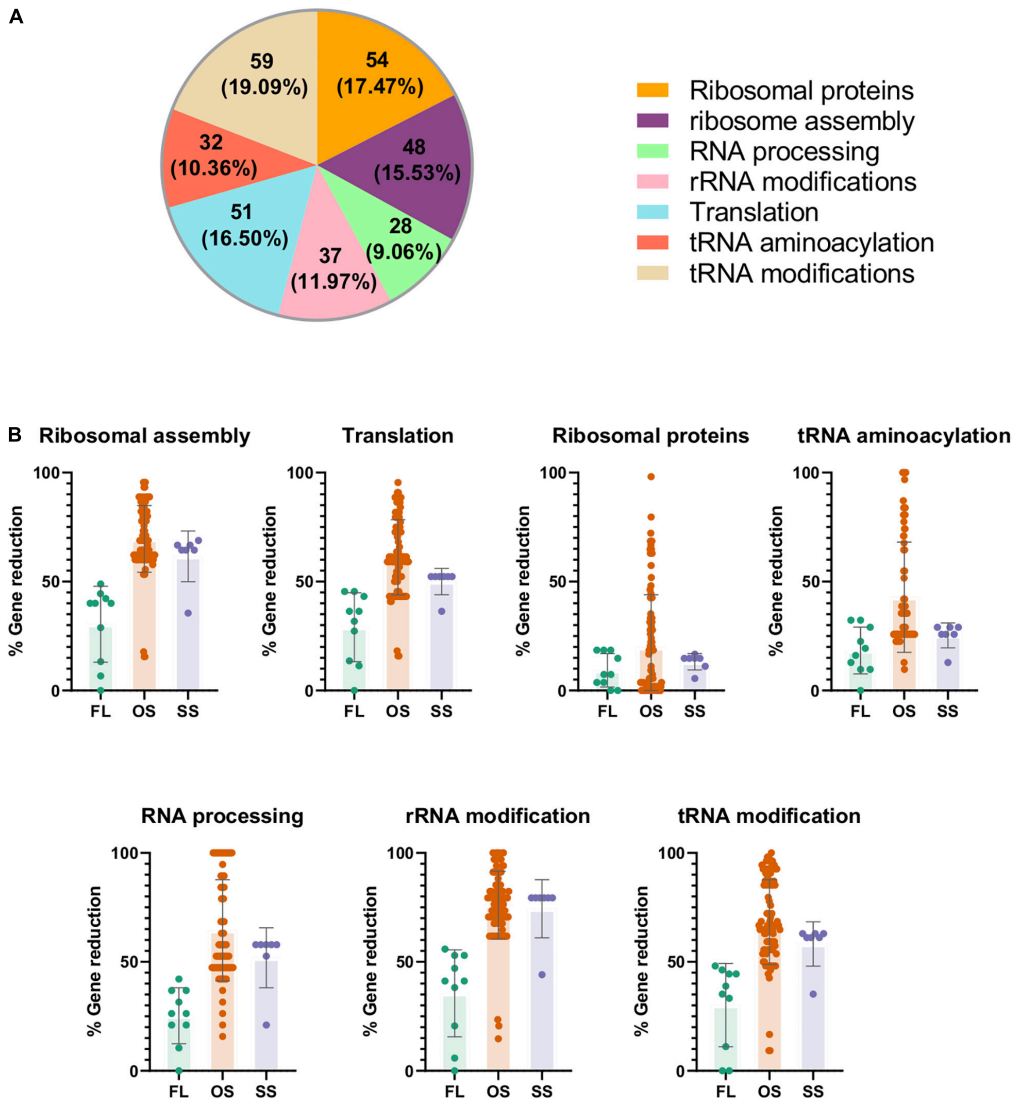


FIGURE 2 | Representation and analysis of the universe of translational genes in the data under study. The complete list of genes can be found in Supplementary Material 2. **(A)** Translational set composed by 309 genes after manual curation. **(B)** Gene reduction level of the genomes under study measured as percentage of translational orthologs not detected in each organism related to the maximum number of genes found in each translational category of the dataset. FL: free-living. OS: obligate endosymbiont. SS: secondary symbiont.

transcription and translation). For this reason, to generate our final genome dataset we considered all coprimary symbionts as a single entity (consortia sheet in **Supplementary Material 1**), leading to only 92 entries. Next, we searched for the orthologs in the genome's dataset using the Roary software. We found 101,012 clusters of orthologs which compose the whole pangenome of the organisms under study.

Computational Search of the Universe of Genes for Translation

Before determining our universe of translational genes, we had to cope with the existence of poorly or wrongly annotated genes and pseudogenes within the analyzed genomes. In fact, a critical difficulty in carrying out this work has been the automation of the process because, even though great efforts are being made to unify the nomenclature (as defined by the International Nucleotide Sequence Database Collaboration, INSDC; Brunak et al., 2002), at present no database provides the unified names for all genes. Most gene descriptions are still based on their initial identification by classical genetics in each given organism, and a recent discovery of their function is often associated to errors, such as entering a function twice without unified descriptors or simply by not including it in databases. For this work, the UniProt nomenclature recommended by the INSDC has been used; in cases where there was no classification, the accepted nomenclature for in *E. coli* K-12 MG1655 (taxonomy ID 511145) was chosen (Schoch et al., 2020).

Once this problem was solved, we set up a *universe of translational protein-coding genes*, which is composed by the genes that encode the proteins that integrate the translation apparatus, as well as those that directly or indirectly participate in the different stages of rRNA or tRNA processing. This gene set was defined based on the genomes of the two selected model bacteria, *E. coli* K-12 MG1655 and *B. subtilis* 168. We detected a total of 309 unique genes (**Supplementary Material 2**). Most of them belong to the ribosomal proteins and tRNA modification categories, and only a few of them to RNA processing (**Figure 2A**).

Then, we compared the previously defined pangenome to the universe of translational genes, to search for the 309 genes involved in translation. As expected, the genome reduction process affects the total number of genes for this function, and the losses depend on the translational subprocess involved (**Figure 2B**). Most ribosomal proteins are present in all organisms regardless of their lifestyle, confirming the importance of the whole ribosome as a functional unit. In contrast, many genes of the other translational categories have been lost, especially in OS and, to a lesser extent, in SS, certainly affected by the genome reduction syndrome. The less conserved genes belong to the ribosome assembly and rRNA/tRNA modification categories, as previously described (de Crécy-Lagard, 2007; Grosjean et al., 2014).

Next, we explored the datasets for essential and persistent genes (Acevedo-Rocha et al., 2013) to get closer to the definition of a functional minimized translation machinery. Evidently, if all organisms have an ortholog of a given gene, its function must be essential and must be included in the minimal translation machinery. Conversely, a gene that only presents orthologs in few organisms with reduced genomes, would not be essential. Based on this notion, a hierarchical clustering analysis was carried out, grouping the data in each of the seven translational subprocesses considered. This machine-learning unsupervised classification algorithm managed to separate each subset of data into two clusters (**Supplementary Figure 1**). A total of 134 orthologs were found in the seven clusters with higher counts and represent the candidate genes to be included in a first proposal of a minimal translational machinery (**Supplementary Material 4**). In previous works, *Buchnera aphidicola BCc* the OS of the aphid *Cinara cedri* (Pérez-Brocal et al., 2006; abbreviated as *bcc* in our study) was considered to possess a small genome close to what could be considered a minimal genome, still able to support the translation process autonomously, while tiny genomes were those that had already lost this ability, so that even genes essential to the process had been lost, making them dependent on the cooperation of a cosymbiont, or even the host, to perform translation (Reyes-Prieto et al., 2014). Therefore, as a first proxy for validation of our approach to define a minimal translational machinery, we compared our results with the genes from the *bcc* genome as a naturally minimized reference genome. The *bcc* genome retains around half of the universe of translational orthologs (150 out of 309 genes), of which it shares 125 genes with our first minimal proposal. Additionally, our proposal contains nine genes (*queA*, *rlmB*, *rne*, *rnhA*, *rplR*, *tgt*, *trmB*, *tsaC*, and *tusE*) that are not present in *bcc*. As *bcc* is cosymbiont of *Serratia symbiotica* str. "Cinara cedri" (SS of *Cinara cedri*; abbreviated *ssz*), we searched for those genes in the cosymbiont's genome. The presence of all nine genes in *ssz* indicates that it probably contributes essential translation genes to the symbiotic relationship. On the other side, the 25 additional genes present in *bcc* and absent in our minimal proposal might reflect that its genome reduction is still an ongoing process. However, as 14 out of these 25 genes are essential in *E. coli* (see next section) and belong to all possible translational subcategories, it cannot be ruled out that some of them are necessary to improve the efficiency of translation in the specific intracellular environment of this bacterium.

Refining the Minimal Translation Machinery

In order to test the viability of our first proposal, we compared it with the set of essential genes involved in translation of *E. coli* K-12 MG1655 and *B. subtilis* 168, according to the DEG database (Luo et al., 2021) and more recent studies on essentiality in *B. subtilis* (Koo et al., 2017; Pedreira et al., 2022) and *E. coli* (Goodall et al., 2020), and with the synthetic JCVI-syn3.0 organism (Hutchison et al., 2016). These two comparisons provide complementary information. While the DEG database reports those genes indispensable for the immediate survival of an organism, the information obtained from JCVI-syn3.0 also highlights the importance of persistent genes, needed for the cell to maintain itself for an extended term. Hutchison et al. (2016) indicate that the JCVI-syn3.0 genome has 195

genes for genetic information storage and processing. Among them, we found that 144 genes are involved in translational processes. Additionally, because 92 genes without assigned biological function were predicted when the JCVI-syn3.0 genome was published (Hutchison et al., 2016), we searched for putative functions of this last set of genes by BLASTP against the non-redundant protein sequences (nr) database from the NCBI web page. Several possibly interesting enzymes were found: a bifunctional oligoribonuclease and PAP phosphatase (*nrnA;* EC 3.1.3.7), a putative pre-16S rRNA nuclease RNaseH-like (*yqgF;* EC 3.1.-.-) and a ribosomal L7Ae/L30e/S12e/Gadd45 family protein (EC 3.1.26.5). The later must be a *rpmD*-like gene, as no *rpmD* has been annotated on the JCVI-syn3.0 genome, although it is present in all minimal sets we are working with, an indication that it must be essential. Moreover, a putative duplicated *pheT* gene was found. **Figure 3** shows the comparison among the three translational datasets. All three datasets, share 95 genes, thus confirming that not all the genes needed are strictly essential (Acevedo-Rocha et al., 2013). Finally, of the 147-genes identified as components of the JCVI-syn3.0 translation machinery, 106 genes (72.1%) are shared with our new proposal for minimal a translation machinery and correspond mainly to ribosomal proteins and aminoacylation enzymes (**Supplementary Material 4**).

Lastly, based on information from the BioCyc and UniProt databases, we performed a functional analysis of the genes found in JCVI-syn3.0 that had not been included in our first proposal, to improve the efficiency of the translational machinery model. Furthermore, through this functional analysis we checked if any function from our first proposal was unnecessary or redundant. For example, in the cases in which there were specific genes of Gram-negative or Gram-positive bacteria (i.e., non-orthologous gene displacement), since this study is based mainly on data from Gram-negative bacteria, we chose to include the alternative corresponding to this group. Thus, regarding aminoacyl-tRNA synthetases (EC 6.1.1.-), our proposal has included *glnS* and *gltX* instead of *gluS*, *gatA*, *gatB*, and *gatC*. As for the tRNA modification genes, we think that only *thiI* and *tsaE* need to be added. Both are related to the modifications at position 8 and 37 of tRNA, respectively. ThiI (EC 2.8.1.4) is an enzyme that produces 4-thiouridine [s(4)U8] (Kambampati and Lauhon, 2000), and it is encoded by an essential gene in *E. coli* K-12 MG1655 (Rajakovich et al., 2012). Moreover, IscS-IscU (EC 2.8.1.7 and EC 3.6.4.10, respectively), two partners of ThiI involved in biological iron–sulfur cluster assembly, needed for sulfur transfer, are already included in our proposal, are essential in both *E. coli* and *B. subtilis*, and are present in JCVI-syn3.0 genome. Finally, we consider that *iscA* should be included as well, although it is not in any of the results shown so far (i.e., JCVI-syn3.0, DEG or our first proposal). This decision is because IscA is necessary for the proper operation of the IscS-IscU system. Although any other alternative enzyme of the HesB family could replace it (López-Madrigal et al., 2013), at least one of them must be part of the minimal translation machinery. On the other hand, TsaE is involved in the formation of a threonylcarbamoyl group (t_6A37), a universally conserved modification (Thiaville et al., 2015). It acts with TsaB, TsaC (EC 2.7.7.87) and TsaD (EC 2.3.1.234), encoded by essential genes included in DEG, and present both in our first proposal and in the JCVI-syn3.0 genome. Regarding to the subcategory translation, we added *tufA* and *tufB* genes, both responsible for the formation of the EF-Tu protein whose essential function is delivering aminoacylated tRNA into the A-site of the ribosome during protein biosynthesis (Kacar et al., 2017). In addition of being necessary for proper translation, they are considered essential in the DEG databases, and are present in the JCVI-syn3.0 genome. As for the subcategory Ribosome assembly, we think that *rimM* must be added because, together with *rbfA,* is needed for efficient processing of 16S rRNA in *E. coli* (Bylund et al., 1998). The *cca* gene (tRNA aminoacylation) which encodes the enzyme Ccase (EC 2.7.7.72) that adds and repairs the 3'-terminal CCA sequence in tRNAs is not included in JCVI-syn3.0. In our results this gene is present to compensate the tendency to have tRNA without the CCA end in OS genomes (data not shown). Finally, nine ribosomal proteins that are present in the JCVI-syn3.0 genome are not include in our first proposal: *rplJ* (L10), *rpmC* (L29), *rpmD* (L30), *rpmE* (L31), *rpmF* (L32), *rpmJ* (L36), *rpsP* (S16), *rpsR* (S18), and *rpsU* (S21). It is known that the composition of the large subunit of the ribosome is less conserved than the one of the small subunit. Moreover, genes of S21, L30 and L31 ribosomal proteins have been consistently reported to be missing (Nikolaeva et al., 2021). Based on these facts, we have only added *rpsP* and *rpsR* to our minimal translation machinery.

The final minimal translation machinery proposed in our study (**Figure 4**) is composed of 142 genes, 113 of which are shared with JCVI-syn3.0 (76.8%), while 112 and 87 (78.8 and 61.3%) are essential in *E. coli* K-12 MG1655 and *B. subtilis* 168, respectively.

DISCUSSION

Based on a computational comparison of genomic information about model and highly reduced bacterial genomes available in public databases plus modern machine learning techniques, we propose a minimal translational gene-set that consists of 142 genes. This work goes beyond the previous proposal of a minimal translation machinery established in Mollicutes (Grosjean et al., 2014), as it includes information about both Gram-positive and Gram-negative organisms with naturally and artificially-reduced genomes, plus a thorough manual curation of all the conjoined information to search for possible mislead or missed genes to define the minimal gene-set implied in a universal translation process. Nevertheless, there is broad agreement with the minimal translational machinery defined by Grosjean et al. (2014).

Ribosomal proteins are part of the ribosome together with rRNAs. In our model, we include 46 genes out of 54 related to this category. Of these, 43 are included in the JCVI-syn3.0 genome and 49 are largely conserved in Mollicutes (Grosjean et al., 2014). It has been experimentally described that at least half of the ribosomal proteins in *E. coli* and *B. subtilis* are not essential for cell survival when individually deleted (Shoji et al., 2011; Akanuma et al., 2012). The absence of some ribosomal proteins in

FIGURE 3 | Comparison of translational genes in DataBases [bacterial gene list of essential translational genes for *Escherichia coli* K-12 MG1655 and *Bacillus subtilis* 168, obtained from the Database of Essential Genes, Subtiwiki, Goodall et al. (2020) and Koo et al. (2017)], cosym (dataset of candidate genes included in our preliminary proposal of a minimal translational machinery obtained from endosymbiont genomes by HCA in this work) and JCVI-syn3.0 (dataset of genes annotated in this genome as involved in translation).

our set can be explained by two non-exclusive reasons. It cannot be discarded that, due to their small size, some apparently absent ribosomal genes are present but have not been detected (i.e., annotated) in the genomes under study. Otherwise, they could have been in fact lost in tiny bacterial genomes, which tend to lose mainly proteins of the large subunit located on the surface of the ribosome (Galperin et al., 2021; Nikolaeva et al., 2021). Regarding this point, seven genes that have not been included in our proposal code for large ribosomal subunit components. Altogether, the ribosomal proteins we have included in our proposal are consistent with the theoretical or experimental ribosomal-proteins sets described in the literature.

Translation factors perform diverse functions throughout the translation process, optimizing it. Most of them were conserved in the 39 Mollicutes genomes studied. Of the 51 genes from the universal set under study that belong to the category Translation, 17 are included in our proposal, and match genes present in the JCVI-syn3.0 genome except for *prfB,* a gene that is absent in most Mollicutes. It codes for release factor 2 (RF-2), the one that recognizes the UAG codon, which is

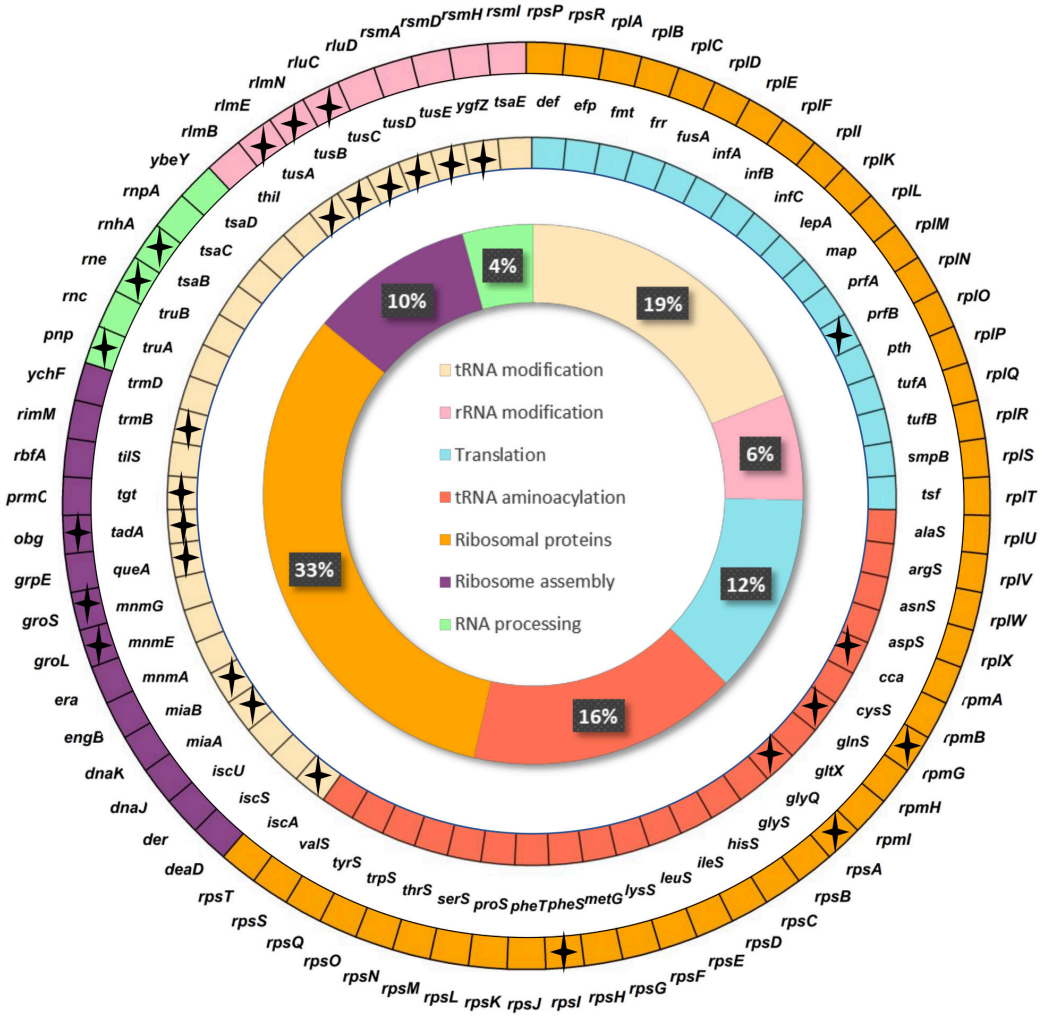


FIGURE 4 | Minimal translational machinery proposed in this work. Genes are classified by translational subfunction. Inner circle shows the percentage of genes in each subcategory based on hierarchical cluster analysis (HCA) analysis. Stars indicate genes not found in JCVI-syn3.0.

recoded from stop to tryptophan in most of these clade members (Grosjean et al., 2014).

Ribonucleases (RNases) that process different RNA precursors are also needed in translation. Only three RNases were found in all Mollicutes genomes analyzed by Grosjean et al. (2014), which indicates that most of them do not need to be included in the minimal translational gene-set. In our analysis, we have found six genes included in the category RNA processing, only three of which are coincident with JCVI- syn3.0, but not with those found in Mollicutes, thus confirming that ribonucleases are not conserved uniformly in the naturally reduced genomes.

Regarding aminoacyl-tRNA synthetases (EC 6.1.1.-), the enzymes that add the corresponding amino acid to the appropriate tRNA, it has been described that only one of them is conserved for each amino acid in small genomes, while tiny genomes do not retain a whole set, probably because some of these enzymes can exert multifunction (i.e., an aminoacyl-tRNA synthetase can load several different amino acids) (Moran and Bennett, 2014) or because they can borrow those from a cosymbiont or the host (Reyes-Prieto et al., 2014). Grosjean and colleagues found 20 aminoacyl-tRNA synthetases were preserved in Mollicutes, some of which are multimeric, and are not coincident to those found in Gram-negative bacteria (Gil et al., 2004). Thus, based on the first *in silico* proposal, taking into account that the *E. coli* model was our default selection in case of divergence, and to warrant an efficient functioning of the system, we have maintained all 21 genes included in this category in previous studies based on endosymbiont genomes (Gil et al., 2004), which include 20 of the 33 genes of this category found in the Mollicutes dataset, plus *cca* and *glyQ*. The latter, which was not included in previous minimal genome proposals (Gil, 2014), encodes the alpha-subunit of glycyl-tRNA synthetase, and was added because both subunits appear to be necessary for its proper functioning in *E. coli* (Ju et al., 2021).

A large part of the genes involved in the post-transcriptional modifications of tRNAs and rRNAs have been lost in small

genomes (Hansen and Moran, 2012). In fact, this is also the category in which greater diversity is observed in Mollicutes (de Crécy-Lagard et al., 2012; Grosjean et al., 2014). Therefore, these genes appear to be dispensable for a minimal translational machinery, even considering the importance of the chemical modification of specific bases in the structure and function of these RNAs. These modifications have been detected in all domains of life (Sergiev et al., 2011) and can be simple (e.g., methylations, thiolations, pseudouridinations) or complex, including the addition of an amino acid (glycine, taurine, threonine, etc.) in tRNAs. Specifically, the modifications of the anticodon domain let the three-dimensional structure of each tRNA to be set for its proper positioning in the ribosome (Agris, 2008), thus facilitating the correct mRNA decoding. The modifications in positions 34 and 37 are fundamental for this purpose, and it is precisely where the greatest variety of modifications are found (Armengod et al., 2014). The minimal translational gene-set defined in Mollicutes revealed that several tRNA modifying enzymes have not been lost due to genome reduction (Grosjean et al., 2014). Yet, the number of genes involved in tRNA modifications decreases drastically along with genome size in obligate endosymbionts, including some of the genes coding for factors involved in these modifications, such as MnmA (EC 2.8.1.13) and the IscS (EC 2.8.1.7), TrmD (EC 2.1.1.228) and TsaC (EC 2.8.1.7) complexes. Little is known to date about tRNAs modification in endosymbionts because, as they cannot be cultured in the laboratory, it is not possible to analyse the state of modification of tRNAs with conventional techniques. The ongoing renewal and improvement of the latest generation of sequencing techniques are making them a valuable tool to deepen in their study (Zhang et al., 2022). Nevertheless, it has been proposed that the tRNAs from endosymbiont such as *B. aphidicola* must have specific changes in the critical bases that would stabilize the structure of the molecule to exercise its translational function despite having A + T rich sequences (Hansen and Moran, 2012). It is currently unknown if these endosymbiont organisms use any of the host's modifying enzymes or if they can use their own modifying enzymes that have not yet been identified, which would improve the structural state of tRNA to optimize its translational function. After adding all these genes for optimal performance, our proposal includes 27 genes for post-transcriptional modifications, 14 of which match the JCVI-syn.3.0 genome, and 10 match the Mollicutes translation machinery proposal.

As for the modifications of rRNAs, they help the translation process by different mechanisms. Many of them are involved in stabilizing the three-dimensional structure of the ribosome, and they are mostly concentrated around the ribosomal catalytic sites. In addition, they promote the interaction with ligands during the translation process and act as checkpoint marks for control of the process (Sergiev et al., 2011). While there are about 36 rRNA modifications in *E. coli*, only 14 have been described in species of the genus *Mycoplasma* (de Crécy-Lagard, 2007). We retrieved nine genes in our proposal, six out of which are present in JCVI-syn3.0.

Since the JCVI-syn3.0 genome has been experimentally minimized and its viability as an organism has been proven, the 113 orthologous genes shared between this genome and our proposal must be essential. The differences between them can be explained by several non-exclusive causes, including non-orthologous gene displacement, adaptation to different environments and protein multifunctionality.

It is widely accepted that the environment strongly influences what would be essential genes for a minimal cell (Koonin, 2003; Gil and Peretó, 2015). Because our main data source are insect symbionts, the environment is very different from that of a free-living cell or one grown in the laboratory under controlled conditions. This could be reflected in the need of different RNA modifying and RNA processing enzymes, where greater differences were found in our comparisons. Remarkably, 16 of the 29 genes included in our first proposal that are not present in JCVI-syn3.0, code for rRNA/tRNA modifying enzymes. These results suggest that the acquisition (or retention) of RNA-modifying enzymes could play an enriching role for bacteria to survive in different environments. From a biotechnological point of view, these alternative enzymes could be tested to build biological systems adapted to specific environmental conditions. In addition, it remains to be determined whether a non-redundant genetic code would make some tRNAs modifications unnecessary.

We have worked with the concept one gene-one function, which is known to be inaccurate. It cannot be ruled out that proteins with low substrate specificity could replace the function of others (e.g., a specific methylase could be able to methylate non-specifically other substrates), while it is known that many proteins can be involved in more than one, sometimes unrelated, functions (moonlighting proteins; Shirafkan et al., 2021). In addition to experimental validation of genome reductions such as the one achieved with *M. mycoides* JCVI-syn3.0, a better delineation of what should be the minimal number of genes necessary to obtain a simplified but still efficient bacterial translational apparatus can be achieved by using machine-learning methods to detect replacements and moonlighting scenarios.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

RG and MG: conceptualization. MG and MR-P: data curation, orthology analysis, and validation. RG: supervision, funding acquisition, and project administration. MG: writing original draft. All authors contributed to the formal analysis and visualization and reviewed, edited, and approved the final manuscript.

FUNDING

This research was funded by European Regional Development Fund (ERDF) and Ministerio de Ciencia, Innovación y Universidades (Spain), grant number PGC2018-099344-B-I00 and Conselleria d'Educació, Generalitat Valenciana (Spain), grant number PROMETEO/2018/133.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.858983/full#supplementary-material

Supplementary Figure 1 | Dendrograms obtained as a result of HCA for cosym dataset. The analyses have been performed for each translational subprocess (legend A) considering clades (legend B).

Supplementary Material 1 | Genomes and cosymbionts of the study (XLSX file). "Genomes" sheet: List of organisms with the abbreviation code used in this study. "Consortia" sheet: list of cosymbionts (symbiotic consortia) used as unique genome entity.

Supplementary Material 2 | Translational genes set (XLSX file). List of genes with translational functions used as universe in this study.

Supplementary Material 3 | Matrix of presence/absence of genes in the 92 genomes of the study (XLSX file). Zero values indicate absence of the gene.

Supplementary Material 4 | Comparison among cosym, *bcc*, *ssz*, JCVI-syn3.0, and DEG datasets (*Escherichia coli* and *Bacillus subtilis*) used in the study (XLSX file). The abbreviation code is the same as in Supplementary Material 1.

REFERENCES

Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W., and Danchin, A. (2013). From essential to persistent genes: A functional approach to constructing synthetic life. *Trends Genet.* 29, 273–279. doi: 10.1016/j.tig.2012.11.001

Agris, P. F. (2008). Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.* 9, 629–635. doi: 10.1038/embor.2008.104

Akanuma, G., Nanamiya, H., Natori, Y., Yano, K., Suzuki, S., Omata, S., et al. (2012). Inactivation of ribosomal protein genes in *Bacillus subtilis* reveals importance of each ribosomal protein for cell proliferation and cell differentiation. *J. Bacteriol.* 194, 6282–6291. doi: 10.1128/JB.01544-12

Armengod, M. E., Meseguer, S., Villarroya, M., Prado, S., Moukadiri, I., Ruiz-Partida, R., et al. (2014). Modification of the wobble uridine in bacterial and mitochondrial tRNAs reading NNA/NNG triplets of 2-codon boxes. *RNA Biol.* 11, 1495–1507. doi: 10.4161/15476286.2014.992269

Brunak, S., Danchin, A., Hattori, M., Nakamura, H., Shinozaki, K., Matise, T., et al. (2002). Nucleotide sequence database policies. *Science* 298, 1333–1333. doi: 10.1126/science.298.5597.1333b

Bylund, G. O., Wipemo, L. C., Lundberg, L. A. C., and Wikström, P. M. (1998). RimM and RbfA are essential for efficient processing of 16S rRNA in *Escherichia coli*. *J. Bacteriol.* 180, 73–82. doi: 10.1128/jb.180.1.73-82.1998

de Crécy-Lagard, V. (2007). Identification of genes encoding tRNA modification enzymes by comparative genomics. *Methods Enzymol.* 425, 153–183. doi: 10.1016/S0076-6879(07)25007-4

de Crécy-Lagard, V., Marck, C., and Grosjean, H. (2012). Decoding in *Candidatus* Riesia pediculicola, close to a minimal tRNA modification set? *Trends Cell Mol. Biol.* 7, 11–34.

Galperin, M. Y., Wolf, Y. I., Garushyants, S. K., Vera Alvarez, R., and Koonin, E. V. (2021). Nonessential ribosomal proteins in bacteria and archaea identified using Clusters of Orthologous Genes. *J. Bacteriol.* 203:e00058–21. doi: 10.1128/JB.00058-21

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi: 10.1126/science.1190719

Gil, R. (2014). "The Minimal Gene-Set Machinery". *Reviews in Cell Biology and Molecular Medicine*,(ed) R. A. Meyers.(Hoboken, NJ: ohn Wiley & Sons). doi: 10.1002/3527600906.mcb.20130079

Gil, R., Latorre, A., and Moya, A. (2010). *Evolution of Prokaryote-Animal Symbiosis from a Genomics Perspective*. Berlin: Springer, 207–233. doi: 10.1007/978-3-642-13615-3_11

Gil, R., and Peretó, J. (2015). Small genomes and the difficulty to define minimal translation and metabolic machineries. *Front. Ecol. Evol.* 3:123. doi: 10.3389/fevo.2015.00123

Gil, R., Silva, F. J., Peretó, J., and Pereto, J. (2004). Determination of the core of a minimal bacterial gene set determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* 68, 518–537. doi: 10.1128/MMBR.68.3.518

Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., et al. (2020). The essential genome of *Escherichia coli* K-12. *mBio* 9, e2096–e2017. doi: 10.1128/mBio.02096-17

Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., Citti, C., et al. (2014). Predicting the minimal translation apparatus: lessons from the reductive evolution of Mollicutes. *PLoS Genet* 10:e1004363. doi: 10.1371/journal.pgen.1004363

Hansen, A. K., and Moran, N. A. (2012). Altered tRNA characteristics and 3' maturation in bacterial symbionts with reduced genomes. *Nucleic Acids Res.* 40, 7870–7884. doi: 10.1093/nar/gks503

Hutchison, C. A. III, Chuang, R.-Y. R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351, aad6253–aad6253. doi: 10.1126/science.aad6253

Ju, Y., Han, L., Chen, B., Luo, Z., Gu, Q., Xu, J., et al. (2021). X-shaped structure of bacterial heterotetrameric tRNA synthetase suggests cryptic prokaryote functions and a rationale for synthetase classifications. *Nucleic Acids Res.* 49, 10106–10119. doi: 10.1093/nar/gkab707

Kacar, B., Garmendia, E., Tuncbag, N., Andersson, D. I., and Hughes, D. (2017). Functional constraints on replacing an essential gene with its ancient and modern homologs. *MBio* 8, e1276–e1217. doi: 10.1128/mBio.01276-17

Kambampati, R., and Lauhon, C. T. (2000). Evidence for the transfer of sulfane sulfur from IscS to ThiI during the in vitro biosynthesis of 4-thiouridine in *Escherichia coli* tRNA. *J. Biol. Chem.* 275, 10727–10730. doi: 10.1074/jbc.275.15.10727

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

Koo, B.-M., Kritikos, G., Farelli, J. D., Todor, H., Tong, K., Kimsey, H., et al. (2017). Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst.* 4, 291–305. doi: 10.1016/j.cels.2016.12.013

Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136. doi: 10.1038/nrmicro751

López-Madrigal, S., Balmand, S., Latorre, A., Heddi, A., Moya, A., and Gil, R. (2013). How does *Tremblaya princeps* get essential proteins from its nested partner *Moranella endobia* in the Mealybug *Planoccocus citri*? *PLoS One* 8:e77307. doi: 10.1371/journal.pone.0077307

Łukasik, P., Nazario, K., Van Leuven, J. T., Campbell, M. A., Meyer, M., Michalik, A., et al. (2018). Multiple origins of interdependent endosymbiotic complexes in a genus of cicadas. *Proc. Natl. Acad. Sci. U.S.A.* 115, E226–E235. doi: 10.1073/pnas.1712321115

Luo, H., Lin, Y., Liu, T., Lai, F. L., Zhang, C. T., Gao, F., et al. (2021). DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Res.* 49, D677–D686. doi: 10.1093/nar/gkaa917

Moran, N. A., and Bennett, G. M. (2014). The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68, 195–215. doi: 10.1146/annurev-micro-091213-112901

Moya, A., Peretó, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9, 218–229. doi: 10.1038/nrg2319

Mushegian, A. (1999). The minimal genome concept. *Curr. Opin. Genet. Dev.* 9, 709–714. doi: 10.1016/S0959-437X(99)00023-4

Nikolaeva, D. D., Gelfand, M. S., and Garushyants, S. K. (2021). Simplification of ribosomes in bacteria with tiny genomes. *Mol. Biol. Evol.* 38, 58–66. doi: 10.1093/molbev/msaa184

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: Rapid large-scale prokaryote pangenome analysis. *Bioinformatics* 31:btv421. doi: 10.1093/bioinformatics/btv421

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Pedreira, T., Elfmann, C., and Stülke, J. (2022). The current state of *Subti*Wiki, the database for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* 50, D875–D882. doi: 10.1093/nar/gkab943

Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., et al. (2006). A small microbial genome: the end of a long symbiotic relationship? *Science* 314, 312–313. doi: 10.1126/science.1130441

Rajakovich, L. J., Tomlinson, J., and Dos Santos, P. C. (2012). Functional analysis of *Bacillus subtilis* genes involved in the biosynthesis of 4-thiouridine in tRNA. *J. Bacteriol.* 194, 4933–4940. doi: 10.1128/JB.00842-12

Reyes-Prieto, M., Latorre, A., and Moya, A. (2014). Scanty microbes, the "symbionelle" concept. *Environ. Microbiol.* 16, 335–338. doi: 10.1111/1462-2920.12220

Reyes-Prieto, M., Vargas-Chávez, C., Llabrés, M., Palmer, P., Latorre, A., and Moya, A. (2020). An update on the Symbiotic Genomes Database (SymGenDB): a collection of metadata, genomic, genetic and protein sequences, orthologs and metabolic networks of symbiotic organisms. *Database* 2020:baz160. doi: 10.1093/database/baz160

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., et al. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* 2020, baaa062. doi: 10.1093/database/baaa062

Schrödinger, L., and DeLano, W. (2020). *PyMOL*. Available at online : http://www.pymol.org/pymol

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sergiev, P. V., Golovina, A. Y., Prokhorova, I. V., Sergeeva, O. V., Osterman, I. A., Nesterchuk, M.l V, et al. (2011). "Modifications of ribosomal RNA: From enzymes to function" in *Ribosomes*:, eds M. V. Rodnina, W. Wintermeyer, and R. Green (Vienna: Springer Vienna), doi: 10.1007/978-3-7091-0215-2

Shirafkan, F., Gharaghani, S., Rahimian, K., Sajedi, R. H., and Zahiri, J. (2021). Moonlighting protein prediction using physico-chemical and evolutional properties via machine learning methods. *BMC Bioinform.* 22:261. doi: 10.1186/s12859-021-04194-5

Shoji, S., Dambacher, C. M., Shajani, Z., Williamson, J. R., and Schultz, P. G. (2011). Systematic chromosomal deletion of bacterial ribosomal protein genes. *J. Mol. Biol.* 413, 751–761. doi: 10.1016/j.jmb.2011.09.004

Silvian, L. F., Wang, J., and Steitz, T. A. (1999). Insights into editing from an ile-tRNA synthetase structure with tRNAile and mupirocin. *Science* 285, 1074–1077. doi: 10.1126/science.285.5430.1074

Sloan, D. B., and Moran, N. A. (2012). Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol. Biol. Evol.* 29, 3781–3792. doi: 10.1093/molbev/mss180

The Blast Sequence Analysis Tool. (2022). *The NCBI Handbook - NCBI Bookshelf*. Available online at: https://www.ncbi.nlm.nih.gov/books/NBK21097/ [Accessed on Jan 8 2022]

Thiaville, P. C., El Yacoubi, B., Köhrer, C., Thiaville, J. J., Deutsch, C., Iwata-Reuyl, D., et al. (2015). Essentiality of threonylcarbamoyladenosine (t^6A), a universal tRNA modification, in bacteria. *Mol. Microbiol.* 98, 1199–1221. doi: 10.1111/mmi.13209

UniProt Consortium, T. U. (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

Wickham, H. (2016). *ggplot2 Elegant Graphics for Data Analysis*. New York: Springer.

Zhang, W., Foo, M., Eran, A. M., and Pan, T. (2022). RNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. *Mol. Cell.* 82, 891–906. doi: 10.1016/j.molcel.2021.12.007

Ziegler, M., and Takors, R. (2019). "Reduced and minimal cell factories in bioprocesses: Towards a streamlined chassis," in *Minimal Cells: Design, Construction, Biotechnological Applications*, ed A. R. Lara, G. Gosset (Berlin: Springer International Publishing), 1–44. doi: 10.1007/978-3-030-31897-0_1

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

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
Contact us: frontiersin.org/about/contact

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