A scanning electron micrograph (SEM) of Escherichia coli bacteria. The bacteria are rod-shaped and appear in various orientations, some individually and some in small clusters. They are set against a dark, textured background. The image is overlaid with a semi-transparent orange band at the top and a semi-transparent green band at the bottom, which serve as backgrounds for the title and publisher information respectively.

EMERGING APPROACHES FOR TYPING, DETECTION, CHARACTERIZATION, AND TRACEBACK OF *ESCHERICHIA COLI*, 2nd EDITION

EDITED BY : Pina M. Fratamico, Chitrita DebRoy and David S. Needleman
PUBLISHED IN : Frontiers in Microbiology



frontiers

Frontiers Copyright Statement

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

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

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

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

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

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

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

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

ISSN 1664-8714

ISBN 978-2-88945-433-4

DOI 10.3389/978-2-88945-433-4

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

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

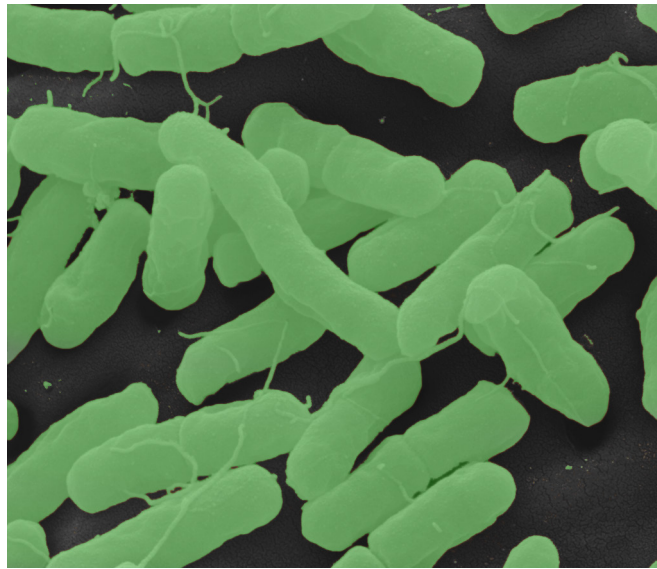
EMERGING APPROACHES FOR TYPING, DETECTION, CHARACTERIZATION, AND TRACEBACK OF *ESCHERICHIA COLI*, 2nd EDITION

Topic Editors:

Pina M. Fratamico, United States Department of Agriculture, Eastern Regional Research Center, USA

Chitrita DebRoy, Pennsylvania State University, USA

David S. Needleman, United States Department of Agriculture, Eastern Regional Research Center, USA



Scanning electron micrograph of Shiga toxin-producing *E. coli* O157:H7 on a lettuce leaf. Image colorized by Susan Reed

Pathogenic *Escherichia coli* strains cause a large number of diseases in humans, including diarrhea, hemorrhagic colitis, hemolytic uremic syndrome, urinary tract infections, and neonatal meningitis, while in animals they cause diseases such as calf scours and mastitis in cattle, post-weaning diarrhea and edema disease in pigs, and peritonitis and airsacculitis in chickens. The different *E. coli* pathotypes are characterized by the presence of specific sets of virulence-related genes. Therefore, it is not surprising that pathogenic *E. coli* constitutes a genetically heterogeneous family of bacteria, and they are continuing to evolve. Rapid and accurate molecular methods are

critically needed to detect and trace pathogenic *E. coli* in food and animals. They are also needed for epidemiological investigations to enhance food safety, as well as animal and human health and to minimize the size and geographical extent of outbreaks. The serotype of *E. coli* strains has traditionally been determined using antisera raised against the >180 different O- (somatic) and 53 H- (flagellar) antigens. However, there are many problems associated with serotyping, including: it is labor-intensive and time consuming; cross reactivity of the antisera with different serogroups occurs; antisera are available only in specialized laboratories; and many strains are non-typeable. Molecular serotyping targeting O-group-specific genes within the *E. coli* O-antigen gene clusters and genes that are involved in encoding for the different flagellar types offers an improved approach for determining the *E. coli* O- and H-groups. Furthermore, molecular serotyping can be coupled with determination of specific sets of virulence genes carried by the strain offering the possibility to determine O-group, pathotype, and the pathogenic potential simultaneously. Sequencing of the O-antigen gene clusters of all of the known O-groups of *E. coli* is now complete, and the sequences have been deposited in the GenBank database. The sequence information has revealed that some *E. coli* serogroups have identical sequences while others have point mutations or insertion sequences and type as different serogroups in serological reactions. There are also a number of other ambiguities in serotyping that need to be resolved. Furthermore, new *E. coli* O-groups are being identified. Therefore, there is an essential need to resolve these issues and to revise the *E. coli* serotype nomenclature based on these findings. There are emerging technologies that can potentially be applied for molecular serotyping and detection and characterization of *E. coli*. On a related topic, the genome sequence of thousands of *E. coli* strains have been deposited in GenBank, and this information is revealing unique markers such as CRISPR (clustered regularly interspaced short palindromic repeats) and virulence gene markers that could be used to identify *E. coli* pathotypes. Whole genome sequencing now provides the opportunity to study the role of horizontal gene transfer in the evolution and emergence of pathogenic *E. coli* strains. Whole genome sequencing approaches are being investigated for genotyping and outbreak investigation for regulatory and public health needs; however, there is a need for establishing bioinformatics pipelines able to handle large amounts of data as we move toward the use of genetic approaches for non-culture-based detection and characterization of *E. coli* and for outbreak investigations.

Citation: Fratamico, P. M., DebRoy, C., Needleman, D. S., eds. (2018). Emerging Approaches for Typing, Detection, Characterization, and Traceback of *Escherichia coli*, 2nd Edition. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-433-4

Table of Contents

06 Editorial: Emerging Approaches for Typing, Detection, Characterization, and Traceback of *Escherichia coli*

Pina M. Fratamico, Chitrita DebRoy and David S. Needleman

Molecular serotyping and subtyping of *E. coli*

09 Advances in Molecular Serotyping and Subtyping of *Escherichia coli*

Pina M. Fratamico, Chitrita DebRoy, Yanhong Liu, David S. Needleman, Gian Marco Baranzoni and Peter Feng

17 Genetic Analysis and Detection of fliC_{H1} and fliC_{H12} Genes Coding for Serologically Closely Related Flagellar Antigens in Human and Animal Pathogenic *Escherichia coli*

Lothar Beutin, Sabine Delannoy and Patrick Fach

27 Rapid, Multiplexed Characterization of Shiga Toxin-Producing *Escherichia coli* (STEC) Isolates Using Suspension Array Technology

John M. Carter, Andrew Lin, Laurie Clotilde and Matthew Lesho

33 Six Novel O Genotypes from Shiga Toxin-Producing *Escherichia coli*

Atsushi Iguchi, Sunao Iyoda, Kazuko Seto, Hironobu Nishii, Makoto Ohnishi, Hirohisa Mekata, Yoshitoshi Ogura and Tetsuya Hayashi

Prevalence, detection, and characterization of pathogenic *E. coli*

41 Characterization of Shiga Toxin Subtypes and Virulence Genes in Porcine Shiga Toxin-Producing *Escherichia coli*

Gian Marco Baranzoni, Pina M. Fratamico, Jayanthi Gangiredla, Isha Patel, Lori K. Bagi, Sabine Delannoy, Patrick Fach, Federica Boccia, Aniello Anastasio and Tiziana Pepe

51 Comparison of Methods for the Enumeration of Enterohemorrhagic *Escherichia coli* from Veal Hides and Carcasses

Brandon E. Luedtke and Joseph M. Bosilevac

62 Detection, Characterization, and Typing of Shiga Toxin-Producing *Escherichia coli*

Brendon D. Parsons, Nathan Zelyas, Byron M. Berenger and Linda Chui

74 From Farm to Table: Follow-Up of Shiga Toxin-Producing *Escherichia coli* Throughout the Pork Production Chain in Argentina

Rocío Colello, María E. Cáceres, María J. Ruiz, Marcelo Sanz, Analía I. Etcheverría and Nora L. Padola

81 Inactivation of Uropathogenic *Escherichia coli* in Ground Chicken Meat Using High Pressure Processing and Gamma Radiation, and in Purge and Chicken Meat Surfaces by Ultraviolet Light

Christopher H. Sommers, O. J. Scullen and Shiowshuh Sheen

87 Surveillance of Extended-Spectrum Beta-Lactamase-Producing *Escherichia coli* in Dairy Cattle Farms in the Nile Delta, Egypt

Sascha D. Braun, Marwa F. E. Ahmed, Hosny El-Adawy, Helmut Hotzel, Ines Engelmann, Daniel Weiß, Stefan Monecke and Ralf Ehricht

Molecular characterization and targets for detection and genetic profiling of pathogenic *E. coli*

101 Development of a High Resolution Virulence Allelic Profiling (HReVAP) Approach Based on the Accessory Genome of *Escherichia coli* to Characterize Shiga-Toxin Producing *E. coli* (STEC)

Valeria Michelacci, Massimiliano Orsini, Arnold Knijn, Sabine Delannoy, Patrick Fach, Alfredo Caprioli and Stefano Morabito

116 Revisiting the STEC Testing Approach: Using *espK* and *espV* to Make Enterohemorrhagic *Escherichia coli* (EHEC) Detection More Reliable in Beef

Sabine Delannoy, Byron D. Chaves, Sarah A. Ison, Hattie E. Webb, Lothar Beutin, José Delaval, Isabelle Billet and Patrick Fach

Whole genome sequencing applications for *E. coli*

126 Implementation of Whole Genome Sequencing (WGS) for Identification and Characterization of Shiga Toxin-Producing *Escherichia coli* (STEC) in the United States

Rebecca L. Lindsey, Hannes Pouseele, Jessica C. Chen, Nancy A. Strockbine and Heather A. Carleton

135 Phylogenetic Analyses of *Shigella* and Enteroinvasive *Escherichia coli* for the Identification of Molecular Epidemiological Markers: Whole-Genome Comparative Analysis Does Not Support Distinct Genera Designation

Emily A. Pettengill, James B. Pettengill and Rachel Binet

146 Whole Genome Sequencing for Public Health Surveillance of Shiga Toxin-Producing *Escherichia coli* Other than Serogroup O157

Marie A. Chattaway, Timothy J. Dallman, Amy Gentle, Michael J. Wright, Sophie E. Long, Philip M. Ashton, Neil T. Perry and Claire Jenkins

150 Whole Genome Sequencing for Genomics-Guided Investigations of *Escherichia coli* O157:H7 Outbreaks

Brigida Rusconi, Fatemeh Sanjar, Sara S. K. Koenig, Mark K. Mammel, Phillip I. Tarr and Mark Eppinger



Editorial: Emerging Approaches for Typing, Detection, Characterization, and Traceback of *Escherichia coli*

Pina M. Fratamico^{1*}, Chitrita DebRoy² and David S. Needleman¹

¹ Agricultural Research Service, United States Department of Agriculture, Eastern Regional Research Center, Wyndmoor, PA, USA, ² Department of Veterinary and Biomedical Sciences, E. coli Reference Center, The Pennsylvania State University, University Park, PA, USA

Keywords: *Escherichia coli*, virulence factors, whole genome sequencing, serotyping, detection, subtyping

Editorial on the Research Topic

Emerging Approaches for Typing, Detection, Characterization, and Traceback of *Escherichia coli*

Commensal *E. coli* inhabit the large intestines of humans and animals and are important in maintaining normal intestinal homeostasis. There are also many groups of disease-causing *E. coli*, including diarrheagenic and extra-intestinal pathogenic *E. coli* (ExPEC). *E. coli* strains have been identified primarily based on their O- and H-antigens, defining the *E. coli* serotype. There are approximately 188 somatic O-antigens, 74 capsular K-antigens, 53 flagellar H-antigens, and greater than 60 fimbrial F-antigens in *E. coli* identified based on antigens that produce an immune response in animals. This research topic consists of articles based on the subject of an international workshop on “Emerging approaches for typing, detection and characterization of *Escherichia coli*” held at the Pennsylvania State University in 2015. This workshop brought together well-known scientists from throughout the world to provide a forum for the exchange of ideas with regard to examining the current serotype classification and nomenclature for *E. coli*, emerging pathotypes, and new technologies and whole genome sequencing (WGS) for detection, characterization, and outbreak investigation. Scientists who presented papers at the conference were affiliated with public health laboratories, regulatory agencies, academic institutions, and industry groups from the U.S., United Kingdom, France, Italy, Canada, Denmark, Germany, and Japan, as well as industry groups working on technologies for characterization, detection, identification, and subtyping *E. coli*. This workshop provided a forum to discuss different concepts and practices for typing *E. coli* based on O- and H-antigens and for characterization of pathotypes. Furthermore, there were discussions on the progress made in the area of WGS as a tool for *E. coli* typing, subtyping, characterization, diagnostics, and outbreak investigation, as well as on the evolution and emergence of highly pathogenic strains (Franz et al., 2014).

A mini review, providing an overview on this research topic is “Advances in molecular serotyping and subtyping of *Escherichia coli*” (Fratamico et al.). Phenotypic methods, such as serotyping for O- and H-antigen determination, biotyping based on biochemical characteristics, or bacteriophage typing have been used for differentiating and characterizing *E. coli* for many years; however, these methods are labor intensive, time consuming, and not always accurate. Recent advances in molecular techniques and DNA sequencing technologies have led to the development of typing and subtyping methods for *E. coli* that are more accurate and have better discriminatory power than phenotypic typing methods. WGS of *E. coli* is allowing more rapid and accurate identification of pathogenic strains and is replacing pulsed-field gel electrophoresis (PFGE), a gold standard for investigating food-borne disease outbreaks (Franz et al., 2014; Joensen et al., 2014).

OPEN ACCESS

Edited by:

Ludmila Chistoserdova,
University of Washington, USA

Reviewed by:

Scott H. Harrison,
North Carolina Agricultural and
Technical State University, USA

*Correspondence:

Pina M. Fratamico
pina.fratamico@ars.usda.gov

Specialty section:

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

Received: 28 September 2016

Accepted: 09 December 2016

Published: 27 December 2016

Citation:

Fratamico PM, DebRoy C and
Needleman DS (2016) Editorial:
Emerging Approaches for Typing,
Detection, Characterization, and
Traceback of *Escherichia coli*.
Front. Microbiol. 7:2089.
doi: 10.3389/fmicb.2016.02089

The necessity to analyze large amounts of data generated from WGS has also led to the development of new and more efficient bioinformatics pipelines. Lindsey et al. described and validated the use of a genotyping plug-in within BioNumerics® v7.5 to provide an accurate and cost-effective single workflow to replace the complex suite of workflows currently used to perform the bioinformatics analyses of WGS data.

Parsons et al. presented a review of various methods and strategies for detection and characterization of Shiga toxin-producing *E. coli* (STEC), including the use of various chromogenic agars, enzyme immunoassays, and qPCR. Their group at the Provincial Laboratory of Public Health in Canada have also utilized WGS, single nucleotide polymorphism (SNP), and k-mer analysis for epidemiological investigations, and they compared these methods to pulsed-field gel electrophoresis and multiple-locus variable number tandem repeat analysis. In agreement with other studies, they found that the sequence-based methods offer higher resolution (Salipante et al., 2015). Use of WGS to distinguish *E. coli* O157:H7 outbreak strains was presented by Rusconi et al. Utilizing customized high-resolution bioinformatics sequence typing approaches, the core genomes, mobilome plasticity, and SNPs were determined. In addition to providing higher strain discriminatory power compared to currently used methods such as PFGE, sequence-based strategies offer the advantages of higher throughput and improved cost-effectiveness. Whole genome sequencing was compared to phenotypic serotyping results on non-O157 STEC isolated from human fecal specimens (Chattaway et al.). They found that most isolates for which the O-group could not be identified by serotyping could be O-typed using WGS data. Furthermore, WGS data provided more accurate *stx*-subtyping results compared to PCR. Thus, WGS provided more reliable results for strain identification and characterization compared to traditional serotyping and PCR, enabling a higher level of strain discrimination and the ability to predict the pathogenic potential.

WGS data on enteric pathogens is shedding light into the relationship between *Shigella* species and *E. coli*. *Shigella* species and enteroinvasive *E. coli* (EIEC) are very similar genetically, and it has been proposed that *Shigella* and EIEC be classified as a single pathovar of *E. coli*. The work of Pettengill et al. based on WGS data and phylogenetic analyses of EIEC and *Shigella* also did not support distinct genera designations for these organisms. They suggested that *Shigella* be classified as EIEC, and common O-antigen designations should be used. They also identified a panel of 404 SNP markers that could be used to discriminate among different phylogenetic clades.

Targeted sequencing in combination with PCR has furthered the characterization and typing of *E. coli*. The sequences of the O-antigen gene clusters of 6 STEC O-groups that were found to be non-typeable using a previously reported *E. coli* O-genotyping PCR system were published by Iguchi et al. They then developed specific PCR assays to identify these novel O-groups that they designated as OgN1, OgN8, OgN9, OgN10, OgN12, and OgN31. Therefore, many O-groups that are not able to be identified by traditional serotyping or other methods may represent new *E. coli* O-groups and will require new O-group designations. For distinguishing the H1 and H12 H-types of *E. coli* that have 97.5%

identity in their *fliC*_{H1} and *fliC*_{H12} genes, Beutin et al. analyzed the sequences of *E. coli* H1 and H12 strains and developed a two-step real-time PCR detection procedure where the first step PCR assay would detect both H1 and H12 flagellar types followed by second step real-time PCR that discriminated H1 or H12. These technologies will ultimately lead to further clarification *E. coli* O- and H-genotypes and in their phylogenetic relationships.

In an effort to develop more accurate and specific methods for STEC detection, Delannoy et al. tested beef enrichments for the presence of specific STEC genetic markers and found that *stx*, *eae*, *espK*, and *espV*, in combination with CRISPR_{O26:H11} served as a suitable set of markers for screening for STEC of concern in beef enrichments, thus lowering the number of samples that required additional testing. A High Resolution Virulence Allelic Profiling (HReVAP) approach to determine allelic variants in genes carried on STEC pathogenicity islands (LEE locus, OI-122, and OI-57) was developed by Michelacci et al. Following cluster analysis of the allelic forms of 91 virulence genes, representing allelic signatures based on HReVAP analyses, they investigated the phylogeny of STEC and identified subpopulations within groups. The approach provided evidence for co-evolution of LEE and OI-122, likely acquired through a single event, and it could be used a tool for studying the evolution of STEC mobile genetic elements. Knowledge of allelic signatures in STEC isolates and the presence of specific genes associated with strains that cause severe illness (hemorrhagic colitis and hemolytic uremic syndrome) improves the ability to test for and identify STEC strains of greater clinical significance.

Knowledge of the virulence genes of food-borne pathogens carried by animals provides information to assess their genetic diversity and zoonotic potential. Baranzoni et al. characterized STEC strains in pigs and found that strains that carried *stx2e* (81%) were the more predominant STEC, followed by strains that carried *stx1a* (14%), *stx2d* (3%), and *stx1c* (1%) as determined by PCR and an *E. coli* Identification (ECID) Array. The swine isolates also carried other virulence genes such as *iha*, *lpfAO26*, *lpfAO157*, *fedA*, *orfA*, and *orfB*. The work presented new insights into the virulence genes associated with porcine isolates that may potentially cause human illness. STEC prevalence throughout the pork production chain in Argentina was studied by Colello et al. by determining the O-groups and virulence gene content of strains collected from farms, during slaughter and processing, and at retail markets. The prevalence and characterization of STEC strains throughout the system reflected vertical transmission of the strains, emphasizing the importance of integrated STEC control systems from farm to table.

Overuse and misuse of antibiotics in humans and animals has led to an increase in antibiotic resistant bacteria, which is a major public health concern worldwide. The prevalence of extended-spectrum beta-lactamase (ESBL)-producing *E. coli* in dairy cattle farms in Egypt was studied by Braun et al. using DNA microarray-based assays that detect resistance genes, and the isolated strains were also serotyped using a SeroGenoTyping array. The study showed a notable prevalence of ESBL-producing strains, and carbapenemase genes (*bla*_{OXA-48} and *bla*_{OXA-181}) (encode enzymes in the beta-lactamase family) were also

detected in isolates resistant to imipenem and meropenem. The high prevalence of antibiotic resistant *E. coli* strains in cattle farms in Egypt points to the importance of increased surveillance efforts in developing countries. Another array-based approach for STEC characterization was reviewed by Carter et al. The “suspension array” technology, which can potentially analyze up to 500 targets in one reaction, was used in both multiplex PCR and immunoassay microbead-based formats to detect *E. coli* serogroup-specific markers, as well as virulence markers. In addition to virulence- and serotype-specific genes, high-throughput array-based systems can potentially detect antibiotic resistance determinants, as well.

Cattle are an important reservoir of STEC, and STEC that cause serious human illness can be found in higher rates from veal calves compared to beef cattle. Methods for determining the prevalence of STEC in veal hides and carcasses, as well as for enumeration of the STEC were evaluated. The methods included a molecular-based most probable number method (MPN), a qPCR assay, and a digital PCR assay (Luedtke and Bosilevac). Each molecular method had strengths and weaknesses related to the detection and enumeration rate and dynamic range for enumeration. ExPEC cause the majority of urinary tract infections and are associated with many other types of extraintestinal infections. Food, particularly poultry, can carry ExPEC, particularly antibiotic-resistant strains that cause human

illness (Manges, 2016). Sommers et al. presented information on interventions for control of uropathogenic *E. coli* (UPEC) in ground chicken meat, purge, and chicken meat surfaces. Using non-thermal processing technologies such as high pressure processing, gamma irradiation, and ultraviolet light (UV-C) they showed significant reduction in the levels of UPEC strains in poultry meat; therefore, these treatments may be useful for the production of safer food products, particularly for at-risk consumers.

In summary, many new technologies have become available to enhance the ability to detect, identify, and characterize *E. coli*. WGS in particular is providing a powerful and expanding range of information to identify targets for development of improved intervention strategies and is being implemented for source tracking and as part of routine surveillance systems (Bergholz et al., 2014). We expect that further developments in WGS and other genomic and molecular technologies will continue to contribute to a greater understanding of the pathogenesis of *E. coli* and ultimately provide better resources for improving public health.

AUTHOR CONTRIBUTIONS

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

REFERENCES

- Bergholz, T. M., Switt, A. I., and Wiedmann, M. (2014). Omics approaches in food safety: fulfilling the promise? *Trends Microbiol.* 22, 275–281. doi: 10.1016/j.tim.2014.01.006
- Franz, E., Delaquis, P., Morabito, S., Beutin, L., Gobius, K., Rasko, D. A., et al. (2014). Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int. J. Food Microbiol.* 187, 57–72. doi: 10.1016/j.ijfoodmicro.2014.07.002
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Manges, A. R. (2016). *Escherichia coli* and urinary tract infections: the role of poultry-meat. *Clin. Microbiol. Infect.* 22, 122–129. doi: 10.1016/j.cmi.2015.11.010
- Salipante, S. J., SenGupta, D. J., Cummings, L. A., Land, T. A., Hoogstraal, D. R., and Cookson, B. T. (2015). Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J. Clin. Microbiol.* 53, 1072–1079. doi: 10.1128/JCM.03385-14

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

Copyright © 2016 Fratamico, DebRoy and Needleman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Advances in Molecular Serotyping and Subtyping of *Escherichia coli*[†]

Pina M. Fratamico^{1*}, Chitrita DebRoy², Yanhong Liu¹, David S. Needleman¹,
Gian Marco Baranzoni¹ and Peter Feng³

¹ Eastern Regional Research Center, Agricultural Research Service, United States Department of Agriculture, Wyndmoor, PA, USA, ² Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA,

³ Division of Microbiology, U.S. Food and Drug Administration, College Park, MD, USA

Escherichia coli plays an important role as a member of the gut microbiota; however, pathogenic strains also exist, including various diarrheagenic *E. coli* pathotypes and extraintestinal pathogenic *E. coli* that cause illness outside of the GI-tract. *E. coli* have traditionally been serotyped using antisera against the ca. 186 O-antigens and 53 H-flagellar antigens. Phenotypic methods, including bacteriophage typing and O- and H- serotyping for differentiating and characterizing *E. coli* have been used for many years; however, these methods are generally time consuming and not always accurate. Advances in next generation sequencing technologies have made it possible to develop genetic-based subtyping and molecular serotyping methods for *E. coli*, which are more discriminatory compared to phenotypic typing methods. Furthermore, whole genome sequencing (WGS) of *E. coli* is replacing established subtyping methods such as pulsed-field gel electrophoresis, providing a major advancement in the ability to investigate food-borne disease outbreaks and for trace-back to sources. A variety of sequence analysis tools and bioinformatic pipelines are being developed to analyze the vast amount of data generated by WGS and to obtain specific information such as O- and H-group determination and the presence of virulence genes and other genetic markers.

Keywords: *Escherichia coli*, molecular serotyping, subtyping, detection, identification, whole genome sequencing, O-group, H-type

OPEN ACCESS

Edited by:

Feng Gao,
Tianjin University, China

Reviewed by:

Andrea Isabel Moreno Switt,
Universidad Andrés Bello, Chile
Seamus Fanning,
University College Dublin, Ireland

*Correspondence:

Pina M. Fratamico
pina.fratamico@ars.usda.gov

Specialty section:

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

Received: 26 February 2016

Accepted: 18 April 2016

Published: 03 May 2016

Citation:

Fratamico PM, DebRoy C, Liu Y,
Needleman DS, Baranzoni GM
and Feng P (2016) Advances
in Molecular Serotyping
and Subtyping of *Escherichia coli*.
Front. Microbiol. 7:644.
doi: 10.3389/fmicb.2016.00644

INTRODUCTION

Escherichia coli strains are commensal organisms that are part of the normal intestinal microflora of humans and other mammals. The traditional method for identifying *E. coli* uses antibodies to test for surface antigens: the O- polysaccharide antigens, flagellar H-antigens, and capsular K-antigens (described below). There are currently ~186 different *E. coli* O-groups and 53 H-types, so serotyping is highly complex. There are also many pathogenic groups of *E. coli* that cause disease in humans and animals, including diarrheagenic *E. coli* and the extra-intestinal pathogenic *E. coli* (ExPEC) that cause illness outside of the GI-tract. Diarrheagenic *E. coli* that cause human illness have been classified based on specific sets of virulence genes they carry and the characteristics of the disease they cause (Kaper et al., 2004). These pathotypes include the enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), Shiga toxin-producing *E. coli* (STEC), diffusely adherent *E. coli* (DEAC), and

[†]Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

adherent invasive *E. coli* (AIEC) that have been associated with Crohn's disease. There are also hybrid pathotypes, including the enteroaggregative hemorrhagic *E. coli* (EAHEC) that carry STEC- and EAEC-associated virulence genes. As an example, EAHEC serotype O104:H4, an EAEC that acquired the phage that carried the Shiga toxin gene of STEC, caused a large outbreak in 2011 associated with illness in over 3800 individuals and 54 deaths (Frank et al., 2011). Certain *E. coli* serotypes are often associated with specific pathotypes, such as STEC O157:H7 and O103:H21 (Kaper et al., 2004) that are important STEC, often referred to as enterohemorrhagic *E. coli* (EHEC). Therefore, pathogenic *E. coli* constitutes a genetically heterogeneous family of bacteria, and they continue to evolve.

Extra-intestinal pathogenic *E. coli* cause illness outside of the gastrointestinal tract, including urinary tract infections, meningitis, pneumonia, septicemia, and other types of infections (Russo and Johnson, 2003; Smith et al., 2007). ExPEC that cause illness in poultry are known as avian pathogenic *E. coli* (APEC). Avian colibacillosis caused by APEC is a major cause of morbidity and mortality associated with economic losses in the poultry industry throughout the world. The human gut is a reservoir for ExPEC that cause human illness. When ExPEC leave the GI tract and infect other parts of the body such as the urinary tract, the blood, or the lungs, illness results (Smith et al., 2007). Animals, particularly, poultry and poultry products (eggs), pork/pigs, and beef/cattle, and also companion animals may carry ExPEC, and thus, these pathogens may be acquired through the food supply, and zoonotic pathogens may also be acquired via contact with animals (Vincent et al., 2010; Nordstrom et al., 2013; Mitchell et al., 2015; Singer, 2015). Investigations of community-acquired UTI and outbreaks of UTI suggested common point sources, such as contaminated food products (Nordstrom et al., 2013). Indeed, high genetic similarity, including antibiotic resistance and virulence gene patterns, between APEC and ExPEC strains causing disease in poultry and humans, respectively, has been observed (Smith et al., 2007; Manges and Johnson, 2012). The ability to differentiate commensal *E. coli* from ExPEC and other pathotypes is important for risk assessment and epidemiological and ecological studies. However, a rapid and reliable typing/identification system or criteria that allows this type of discrimination and that also provides information on the organism's evolutionary history, fitness, and pathogenic potential has not yet been established. Determining whether an *E. coli* strain is an ExPEC and whether it is pathogenic is based on its source, O:K:H serotype, phylogenetic background, virulence factor profile, and experimental virulence in an animal model. ExPEC belong to specific phylogenetic groups (A, B1, B2, and D) determined based on multilocus enzyme electrophoresis, ribotyping, or by triplex PCR targeting the genes *chuA* and *yjaA* and a particular DNA fragment known as TSPE4.C2. ExPEC strains belonging to phylogenetic groups B2 and D show higher virulence in humans (Clermont et al., 2000; Smith et al., 2007). It has become evident that certain ExPEC lineages or clonal groups are responsible for a large fraction of human extraintestinal *E. coli* infections, and these lineages are becoming increasingly multi-drug resistant (Smith et al., 2007; Manges and Johnson, 2012).

Rapid and accurate molecular methods are critically needed to detect and trace pathogenic *E. coli* in food and animals and for epidemiological investigations to enhance food safety and animal and human health, as well as to minimize the size and geographical extent of outbreaks. As opposed to traditional serotyping using antisera raised against the different *E. coli* O- and H-types, molecular serotyping generally refers to genetic-based assays targeting O-group-specific genes found within the *E. coli* O-antigen gene clusters and the H-antigen genes that encode for the different flagellar types. Although determining the *E. coli* serotype could be considered a component of subtyping (differentiation beyond the species level), methods used for molecular subtyping such as pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST), and whole genome sequencing (WGS) generate a unique "fingerprint" of the bacterium that can be used in outbreak investigations and to determine the source of illnesses. There are many problems associated with traditional serotyping for determining the *E. coli* O- and H-groups. It is costly, labor-intensive and time consuming, cross reactivity of the antisera with different serogroups occurs, antisera are available only in specialized laboratories, batch-to-batch variations in antibodies can occur, and many *E. coli* strains isolated from various sources are non-typeable (Lacher et al., 2014). Thus, molecular serotyping offers alternative methods for *E. coli* serotyping, and furthermore, they can be coupled with assays for specific virulence gene enabling the determination of O- and H-group, pathotype, and the strain's pathogenic potential simultaneously.

E. coli O-, K-, AND H-ANTIGENS

The outer membrane of *E. coli* is composed of lipopolysaccharides (LPS) that includes lipid A, core oligosaccharides, and a unique polysaccharide, referred to as the O-antigen. Loss of the O-antigens results in attenuated virulence suggesting their importance in host-pathogen interactions (Sarkar et al., 2014). Based on the antigenic diversity among the different O-antigens, they have been targeted as biomarkers for classification of *E. coli* since the 1940s (Kaufmann, 1943, 1944, 1947). Later, Ørskov et al. (1977) presented a comprehensive serotyping system for 164 *E. coli* O-groups and developed a typing scheme based on the presence of three principal surface antigens, O-antigens, flagellar H-antigens, and capsular K-antigens. Since few laboratories had capabilities to type the K antigen, serotyping based on O- and H-antigens became the gold standard for *E. coli* typing. Currently, O-groups numbered O1-O188 have been defined, except for O31, O47, O67, O72, O94, and O122 that have not been designated (Ørskov and Ørskov, 1984; Scheut et al., 2004), and four groups have been divided into subtypes O18ab/ac, O28ab/ac, O112ab/ac, and O125ab/ac, giving a total of 186 O-groups.

The conventional serotyping method is based on agglutination reactions of the O-antigen with antisera that are generated in rabbits against each of the O-groups (Ørskov and Ørskov, 1984). The method is easy to carry out; however, it is

laborious and error-prone, and thus, molecular methods are better alternatives for O-typing (Ballmer et al., 2007; Lacher et al., 2014). The genes that encode for O-antigens are located on the chromosome in a cluster designated as the O-antigen gene cluster (O-AGC). These are flanked by two conserved sequences called JUMPstart, a 39 bp-element at the 5' end (Hobbs and Reeves, 1994), which is downstream of *galF* (UTP-glucose-1-phosphate uridylyltransferase) and *gnd* (6-phosphogluconate dehydrogenase) at the 3' end. Analysis of the O-AGCs of all *E. coli* O-groups (Iguchi et al., 2015a; DebRoy et al., 2016) showed that the sizes of the O-AGCs and their gene content vary considerably, which results in the variability of O-antigens. O-antigens are composed of 10–25 repeating units of two to seven sugar residues and are processed by three mechanism of which the most common is Wzy (O antigen polymerase) dependent, followed by an ABC transporter dependent system, and the third mechanism, which involves a synthase dependent pathway (Greenfield and Whitfield, 2012) by which the O-antigens are flipped across the outer membrane. The pathways for biosynthesis of the O-AGCs and assembly of O-antigens have been studied extensively (Samuel and Reeves, 2003). Each of the O-antigens that utilize Wzy-dependent pathway carries two unique genes *wzx* (O-antigen flippase) and *wzy* (O-antigen polymerase). Wzx proteins translocates the O-units across the inner membrane, and Wzy polymerizes the O-antigen (Samuel and Reeves, 2003). For the ABC transporter-dependent pathway, *wzm* (O-antigen ABC transporter permease gene) and *wzt* (ABC transporter ATP-binding gene) are involved in O-AGC synthesis. The O-AGCs are composed of nucleotide sugar biosynthesis genes that are involved in the synthesis of O-antigen nucleotide sugar precursors, the glycosyl transferases that transfer the various sugar precursors to form the oligosaccharide, and the O-antigen processing genes described above.

All of the O-AGC clusters have been sequenced, and sequence analyses revealed that some O-AGCs are 98–100% identical (Iguchi et al., 2015a; DebRoy et al., 2016) while others have point mutations or insertion sequences which causes these to type as different serogroups (Liu et al., 2008, 2015). Therefore, there is a need to resolve these discrepancies, merge or eliminate serogroups and to revise the *E. coli* serotype nomenclature (DebRoy et al., 2016). Furthermore, many of the *E. coli* O-AGCs have been found to be identical to those of other Enterobacteriaceae members such as *Shigella* and *Salmonella* (Wang et al., 2007). Out of 34 distinct *Shigella* O-antigens, 13 were unique to *Shigella*; however, the other 21 were also found in *E. coli* (Liu et al., 2008). Similarly, out of 46 O-AGCs of *Salmonella*, 24 of were found to be identical or closely related to *E. coli* O-antigens (Liu et al., 2014).

Serology has defined 53 H-flagellar antigens (Ørskov and Ørskov, 1984; Ewing, 1986) that are numbered from H1 to H56, but H-types 13, 22, and 50 are not in use (Ørskov et al., 1975; Centers for Disease Control and Prevention [CDC], 1999). Molecular H-typing methods are based on the sequences of *fliC* gene that encode for the FliC, the flagellar filament structural protein (Wang et al., 2003). The N- and C-terminals of FliC are highly conserved, so different H-types are due to amino

acid differences within the central region, which is the surface-exposed antigenic part of the flagellar filament (Namba et al., 1989). Thus, PCR methods developed to distinguish H-types target the variable region of the *fliC* gene (Machado et al., 2000); however, these regions of some H-types such as H1 and H12 and H25 and H28 are very similar, making them difficult to distinguish. However, a two-step PCR method was developed that can distinguish between *fliC*_{H1} and *fliC*_{H12} (Beutin et al., 2015, 2016). Other methods such as Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF)-based peptide mass fingerprinting in conjunction with a custom *E. coli* H-antigen data base (Cheng et al., 2014) has been also utilized to distinguish H-types (Chui et al., 2015).

METHODS USED FOR SUBTYPING AND MOLECULAR SEROTYPING OF *E. coli*

Subtyping methods that allow for differentiation of *E. coli* beyond the species and subspecies level are critical for determining the source of outbreaks and establishing transmission pathways (Eppinger et al., 2011; Frank et al., 2011). Several phenotype-based and genotype-based methods for subtyping *E. coli* are listed in Table 1. Phenotypic culture methods, in conjunction with biochemical-based testing, serotyping, phage typing, multilocus enzyme electrophoresis have been used for many years and could be considered gold standard methods; however, they are time and labor intensive and may not be very discriminatory.

Compared to phenotypic methods, genetic subtyping methods that are based on bacterial DNA, generally have better discriminatory ability. Of the various methods used for *E. coli* subtyping, PFGE is a reliable and highly discriminating method and has been considered to be the “gold standard” of typing methods. Through the establishment of PulseNet (Ribot et al., 2006), use of PFGE has had a major impact on pathogen subtyping and outbreak investigation.

In contrast to traditional serotyping, Luminex®-based suspension assays allow for simultaneous testing for multiple serogroups in a single assay. Lin et al. (2011) performed PCR assays targeting the *wzx* and *wzy* genes of ten Shiga toxin-producing *E. coli* (STEC) serogroups, and then used the Luminex® system to identify the 10 serogroups through binding of the PCR products to fluorescent microspheres conjugated to specific DNA probes for each of the ten serogroups. Clotilde et al. (2015) used the Luminex® technology, both antibody- and multiplex PCR-based, and compared them to traditional *E. coli* serotyping. The results of the two Luminex® assays were mostly consistent, and 11 STEC isolates that were previously untypeable by traditional serotyping were able to be typed.

Multiplex PCR-based assays targeting unique regions within the *E. coli* O-AGCs have been used to determine the O-groups. A review by DebRoy et al. (2011) describes many of these assays, most of which target the *E. coli* *wzx* and *wzy* genes. Based on O-AGC sequence data for all O-groups, Iguchi et al. (2015b) designed 162 PCR primer pairs for identification and classification of *E. coli* O-serogroups. The primer pairs were

TABLE 1 | Phenotype- and genotype-based methods for subtyping and molecular serotyping of *E. coli*.

Phenotype-based methods	O-antigen ^a	H-antigen ^a	Virulence-related and other genes	SNPs and other markers	Reference
Immunological O/H typing (serotyping)	X	X			Ørskov et al., 1977
Bacteriophage typing				X	Ahmed et al., 1987; Krause et al., 1996
Multilocus enzyme electrophoresis (MLEE)				X	Ochman and Selander, 1984; Selander et al., 1986; Campos et al., 1994
MALDI-TOF		X		X	Chui et al., 2015
Genotype-based methods					
Restriction length polymorphism (RFLP)	X	X	X		Coimbra et al., 2000; Beutin et al., 2005; Moreno et al., 2006; Abbadi and Strockbine, 2007
Luminex-based suspension assay	X				Lin et al., 2011; Clotilde et al., 2015
Amplified fragment length polymorphisms (AFLP)				X	Hahm et al., 2003; Leung et al., 2004
Optical mapping			X	X	Kotewicz et al., 2007, 2008; Miller, 2013
Ribotyping				X	Martin et al., 1996; Carson et al., 2001
Multilocus variable number tandem repeat analysis (MLVA)				X	Hyttiä-Trees et al., 2006; Byrne et al., 2014
Pulsed-field gel electrophoresis			X	X	Krause et al., 1996; Ribot et al., 2006
Multilocus sequence typing (MLST)				X	Eichhorn et al., 2015; Manges et al., 2015
High throughput real-time PCR	X	X	X		Bugarel et al., 2010a,b, 2011a,b; Delannoy et al., 2013; Tseng et al., 2014
Multiplex PCR	X	X	X		Fratamico and DebRoy, 2010; Botkin et al., 2012; Doumith et al., 2012; Fraticamo et al., 2014; Iguchi et al., 2015b
Whole genome sequencing and SNP analysis	X	X	X	X	Zhang et al., 2006; Eppinger et al., 2011; Norman et al., 2012, 2015; Joensen et al., 2014, 2015; Griffing et al., 2015; DebRoy et al., 2016; Ison et al., 2016
Virulence gene profiles				X	Nandanwar et al., 2014; Manges et al., 2015
CRISPRs				X	Shariat and Dudley, 2014; Delannoy et al., 2015
Microarray	X	X	X	X	Liu and Fraticamo, 2006; Hegde et al., 2013; Lacher et al., 2014
NeoSEEK™ (PCR-mass spectroscopy)	X	X	X	X	Stromberg et al., 2015

^aThe O-somatic and H-flagellar antigens define the *E. coli* serotype. Agglutination assays using antibodies that react with the specific O- and H-antigens are the basis for traditional serotyping. Molecular serotyping methods are generally based on genetic targets specific to the O- and H-antigens.

used in 20 separate multiplex PCR assays with each assay containing 6–9 primer pairs that amplified products of different sizes so that they could be distinguished. A high-throughput PCR method based on the GeneDisc® array targeted virulence genes and O- and H-type-specific genes for identification of STEC associated with severe illness (Bugarel et al., 2010b). Another high-throughput method, known as the BioMark™ real-time PCR system (Fluidigm), used a panel of virulence genes as discriminative markers to differentiate EHEC O26 strains,

EHEC-like O26 pathogenic strains, and avirulent O26 strains (Bugarel et al., 2011a).

Clustered regularly interspaced short palindromic repeats (CRISPR) are short, highly conserved DNA repeats separated by unique sequences of similar length, and they have been used for subtyping, identification, and detection of bacteria (Shariat and Dudley, 2014). Based on spacer content or sequencing of CRISPR loci, CRISPR-based typing analyses can be used to differentiate strains for epidemiological investigations or

for detection. Delannoy et al. (2012) utilized CRISPR loci of seven important EHEC serotypes to develop real-time PCR assays, generating results based on CRISPR polymorphisms that correlated with specific EHEC O:H serotypes and the presence of EHEC virulence genes.

DNA microarrays have also been developed for molecular serotyping of *E. coli* (Liu and Fratamico, 2006; Ballmer et al., 2007; Geue et al., 2014; Lacher et al., 2014). One microarray method to identify *E. coli* serogroups involved spotting O-group-specific *wzx* or *wzy* gene oligonucleotides or PCR products onto the chip and hybridized with labeled PCR products of the entire O-AGCs (Liu and Fratamico, 2006). Lacher et al. (2014) reported on the use of an FDA-ECID (*E. coli* identification) microarray for O- and H-typing of *E. coli*. The ECID chip was designed based on >250 *E. coli* genomes and incorporates over 40,000 *E. coli* genes, including O- and H-group-specific genes, and approximately 9800 single nucleotide polymorphisms (SNPs). Antibody-based microarrays have also been developed to detect important non-O157 STEC serogroups (Gehring et al., 2013; Hegde et al., 2013). Although this method is rapid and has the potential to be used for high throughput screening, the utilization of this method is dependent on the availability of antibodies with good specificity.

The commercial introduction of next-generation sequencing technologies has made it possible to perform routine WGS of *E. coli* and other bacteria relatively rapidly and at affordable costs (Franz et al., 2014). Since WGS typing has discriminatory power superior to other typing methods, it has the potential to revolutionize bacterial subtyping. A MLST webserver was designed to determine sequence types (STs) of bacteria using WGS data. STs were determined from uploaded preassembled complete or partial genome sequences or short sequence reads obtained from different sequencing platforms (Larsen et al., 2012). Based on SNPs observed from WGS data, Norman et al. (2015) identified unique STEC O26 genotypes in human and cattle strains. These isolates had similar virulence gene profiles and did not cluster in separate polymorphism-derived genotypes, and thus human and cattle strains could not be distinguished within the phylogenetic clusters. An approach based on targeted amplicon sequencing for SNP genotyping was used to determine the relationship of *stx*-positive and *stx*-negative *E. coli* O26:H11 strains from cattle compared to the genomes of human clinical isolates (Ison et al., 2016). Joensen et al. (2015) described SerotypeFinder, a publicly available web tool hosted by the Center for Genomic Epidemiology, Denmark, which enables WGS-based serotyping of *E. coli*. Typing is based on *wzx*, *wzy*, *wzm*, and *wzt*, as well as flagellin-associated genes. Similar to SerotypeFinder, the VirulenceFinder tool can be used to determine virulence genes in *E. coli* to determine different pathogenic groups (Joensen et al., 2014).

Whole genome sequencing typing has the potential to be the new “gold-standard” for pathogen subtyping. However,

some challenges need to be addressed before standardization and full implementation of this technology. The bioinformatic analyses required to analyze enormous amounts of sequence data generated by WGS are necessitating the development of analysis pipelines to enhance the assembly, annotation, and interpretation of the data, which will require a coordinated international approach (Franz et al., 2014; Oulas et al., 2015). Currently, the following databases for WGS and advanced detection are available: the 100K Genome Project¹, GenomeTrakr Network², Global Microbial Identifier³, and Advanced Molecular Detection⁴. These databases are creating a vast resource of microbial genome information for WGS-based surveillance of microbial pathogens. Furthermore, detailed analysis of WGS data can determine the *E. coli* O- and H-type and provide information on the resistome (antibiotic resistance gene profile) of the isolate, and the presence of specific virulence genes, prophages, and plasmids, as well as other genetic information important to identify *E. coli* pathotypes as well as utility in evolutionary studies. The advantages of WGS approaches are being recognized by academic, government, industry, and the private sector for addressing regulatory and public health needs. However, as we move toward the use of these genetic approaches for non-culture-based detection, characterization, subtyping, trace backs, and outbreak investigations, it will be critical to establish bioinformatics pipelines that are capable of analyzing and handling the large amounts of data that are generated.

AUTHOR CONTRIBUTIONS

PF, CD, YL, DN, GB, and PF have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by an appointment to the Agricultural Research services (ARS) Research Participation Program which is administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the USDA. ORISE is managed by ORAU under DOE contract number DE-AC05-06OR23100. All opinions expressed in this manuscript are the author's and do not necessarily reflect the policies and views of USDA, ARS, DOE, or ORAU/ORISE.

¹ <http://100kgenome.vetmed.ucdavis.edu/>

² <http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/>

³ <http://www.cdc.gov/amd/project-summaries/index.html>

⁴ <http://www.globalmicrobialidentifier.org/>

REFERENCES

- Abbadi, S. H., and Strockbine, N. A. (2007). Identification of *Escherichia coli* flagellar types by restriction of the amplified fliC gene. *Egypt. J. Med. Microbiol.* 16, 225–233.
- Ahmed, R., Bopp, C., Borczyk, A., and Kasatiya, S. (1987). Phage-typing scheme for *Escherichia coli* O157:H7. *J. Infect. Dis.* 155, 806–809. doi: 10.1093/infdis/155.4.806
- Ballmer, K., Korczak, B. M., Kuhnert, P., Slickers, P., Ehricht, R., and Hächler, H. (2007). Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. *J. Clin. Microbiol.* 45, 370–379. doi: 10.1128/JCM.01361-06
- Beutin, L., Delannoy, S., and Fach, P. (2015). Sequence variations in the flagellar antigen genes fliCH25 and fliCH28 of *Escherichia coli* and their use in identification and characterization of enterohemorrhagic *E. coli* (EHEC) O145:H25 and O145:H28. *PLoS ONE* 10:e0126749. doi: 10.1371/journal.pone.0126749
- Beutin, L., Delannoy, S., and Fach, P. (2016). Genetic analysis and detection of fliCH1 and fliCH12 genes coding for serologically closely related flagellar antigens in human and animal pathogenic *Escherichia coli*. *Front. Microbiol.* 7:135. doi: 10.3389/fmicb.2016.00135
- Beutin, L., Strauch, E., Zimmermann, S., Kaulfuss, S., Schaudinn, C., Männel, A., et al. (2005). Genetical and functional investigation of fliC genes encoding flagellar serotype H4 in wildtype strains of *Escherichia coli* and in a laboratory *E. coli* K-12 strain expressing flagellar antigen type H48. *BMC Microbiol.* 5:4. doi: 10.1186/1471-2180-5-4
- Botkin, D. J., Galli, L., Sankarapani, V., Soler, M., Rivas, M., and Torres, A. G. (2012). Development of a multiplex PR assay for detection of Shiga toxin-producing *Escherichia coli*, enterohemorrhagic *E. coli*, and enteropathogenic *E. coli* strains. *Front. Cell Infect. Microbiol.* 2:8. doi: 10.3389/fcimb.2012.00008
- Bugarel, M., Beutin, L., and Fach, P. (2010a). Low-density microarray targeting non-locus of enterocyte effacement effectors (nle genes) and major virulence factors of Shiga toxin-producing *Escherichia coli* (STEC): a new approach for molecular risk assessment of STEC isolates. *Appl. Environ. Microbiol.* 76, 203–211. doi: 10.1128/AEM.01921-09
- Bugarel, M., Beutin, L., Martin, A., Gill, A., and Fach, P. (2010b). Microarray for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with hemorrhagic colitis and hemolytic uremic syndrome in humans. *Int. J. Food Microbiol.* 142, 318–329. doi: 10.1016/j.ijfoodmicro.2010.07.010
- Bugarel, M., Beutin, L., Scheutz, F., Loukiadis, E., and Fach, P. (2011a). Identification of genetic markers for differentiation of Shiga toxin-producing, enteropathogenic and avirulent strains of *Escherichia coli* O26. *Appl. Environ. Microbiol.* 77, 2275–2281. doi: 10.1128/AEM.02832-10
- Bugarel, M., Martin, A., Fach, P., and Beutin, L. (2011b). Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis of molecular risk assessment of typical and atypical EPEC strains. *BMC Microbiol.* 11:142. doi: 10.1186/1471-2180-11-142
- Byrne, L., Elson, R., Dallman, T. J., Perry, N., Ashton, P., Wain, J., et al. (2014). Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 as a routine public health tool in England. *PLoS ONE* 9:e85901. doi: 10.1371/journal.pone.0085901
- Campos, L. C., Whittam, T. S., Gomes, T. A. T., Andrade, J. R. C., and Trabulsi, L. R. (1994). *Escherichia coli* serogroup O111 includes several clones of diarrheagenic strains with different virulence properties. *Infect. Immun.* 62, 3282–3288.
- Carson, C. A., Shear, B. L., Ellersieck, M. R., and Asfaw, A. (2001). Identification of fecal *Escherichia coli* from humans and animals by ribotyping. *Appl. Environ. Microbiol.* 67, 1503–1507. doi: 10.1128/AEM.67.4.1503-1507.2001
- Centers for Disease Control and Prevention [CDC] (1999). *Laboratory Methods for the Diagnosis of Epidemic Dysentery and Cholera*. Available at: <http://www.cdc.gov/cholera/pdf/Laboratory-Methods-for-the-Diagnosis-of-Epidemic-Dysentery-and-Cholera.pdf> [Accessed 14 February, 2016].
- Cheng, K., Sloan, A., McCorrister, S., Babiuk, S., Bowden, T. R., Wang, G., et al. (2014). Fit-for-purpose curated database application in mass spectrometry-based targeted protein identification and validation. *BMC Res.* 7:444. doi: 10.1186/1756-0500-7-444
- Chui, H., Chan, M., Hernandez, D., Chong, P., McCorrister, S., Robinson, A., et al. (2015). Rapid, sensitive and specific *E. coli* H antigen typing by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)-based peptide mass fingerprinting. *J. Clin. Microbiol.* 53, 2480–2485. doi: 10.1128/JCM.00593-15
- Clermont, O., Bonacorsi, S., and Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* 66, 4555–4558. doi: 10.1128/AEM.66.10.4555-4558.2000
- Clotilde, L. M., Salvador, A., Bernard, C. I. V., Lin, A., Lauzon, C., Muldoon, M., et al. (2015). Comparison of multiplex immunochemical and molecular serotyping methods for Shiga toxin-producing *Escherichia coli*. *Foodborne Pathog. Dis.* 12, 118–121. doi: 10.1089/fpd.2014.1827
- Coimbra, R. S., Grimont, F., Lenormand, P., Burguière, Beutin, L., and Grimont, P. A. D. (2000). Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfp-RFLP). *Res. Microbiol.* 151, 639–654. doi: 10.1016/S0923-2508(00)00134-0
- DebRoy, C., Frataamico, P. M., Yan, X., Baranzoni, G. M., Liu, Y., Needleman, D. S., et al. (2016). Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-typing. *PLoS ONE* 11:e0147434. doi: 10.1371/journal.pone.0147434
- DebRoy, C., Roberts, E., and Frataamico, P. M. (2011). Detection of O antigens in *Escherichia coli*. *Anim. Health Res. Rev.* 12, 169–185. doi: 10.1017/S1466252311000193
- Delannoy, S., Beutin, L., and Fach, P. (2013). Discrimination of enterohemorrhagic *Escherichia coli* (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. *J. Clin. Microbiol.* 51, 3257–3262. doi: 10.1128/JCM.01471-13
- Delannoy, S., Beutin, L., and Fach, P. (2015). Improved traceability of Shiga-toxin-producing *Escherichia coli* using CRISPRs for detection and typing. *Environ. Sci. Pollut. Res. Int.* 1–12. doi: 10.1007/s11356-015-5446-y
- Delannoy, S., Beutin, R., and Fach, P. (2012). Use of clustered regularly interspaced short palindromic repeat sequence polymorphism for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. *J. Clin. Microbiol.* 50, 4035–4040.
- Doumith, M., Day, M. J., Hope, R., Wain, J., and Woodford, N. (2012). Improved multiplex PCR strategy for rapid assignment of the four major *Escherichia coli* phylogenetic groups. *J. Clin. Microbiol.* 50, 3108–3110. doi: 10.1128/JCM.01468-12
- Eichhorn, I., Heidemanns, K., Semmler, T., Kinnemann, B., Mellmann, A., Harmsen, D., et al. (2015). Highly virulent non-O157 enterohemorrhagic *Escherichia coli* (EHEC) serotypes reflect similar phylogenetic lineages, providing new insights into the evolution of EHEC. *Appl. Environ. Microbiol.* 81, 7041–7047. doi: 10.1128/AEM.01921-15
- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011). Genetic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20142–20147. doi: 10.1073/pnas.1107176108
- Ewing, W. H. (1986). Edwards and Ewing's Identification of the Enterobacteriaceae. *Int. J. Syst. Evol. Microbiol.* 36, 581–582.
- Frank, C., Werber, D., Jakob, C. J., Askar, M., Faber, M., der Heiden, M., et al. (2011). Epidemic profile of Shiga toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* 365, 1771–1780.
- Franz, E., Delaquis, P., Morabito, S., Beutin, L., Gobius, K., Rasko, D. A., et al. (2014). Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int. J. Food Microbiol.* 187, 57–72. doi: 10.1016/j.ijfoodmicro.2014.07.002
- Fratamico, P. M., and DebRoy, C. (2010). Detection of *Escherichia coli* O157:H7 in food using real-time multiplex PCR assays targeting the stx1, stx2, wzyO157, and the fliCh7 or eae genes. *Food Anal. Methods* 3, 330–337. doi: 10.1007/s12161-010-9140-x
- Fratamico, P. M., Wasilenko, J. L., Garman, B., DeMarco, D. R., Varkey, S., Jensen, M., et al. (2014). Evaluation of a multiplex PCR real-time PCR method for detecting Shiga toxin-producing *Escherichia coli* in beef and comparison to the U.S. Department of Agriculture Food Safety and Inspection Service Microbiology Laboratory Guidebook Method. *J. Food Prot.* 77, 180–188. doi: 10.4315/0362-028X.JFP-13-248
- Gehring, A., Barnett, C., Chu, T., DebRoy, C., D'Souza, D., Eaker, S., et al. (2013). A high-throughput antibody-based microarray typing platform. *Sensors (Basel)* 13, 5737–5748. doi: 10.3390/s130505737

- Geue, L., Monecke, S., Engelmann, I., Braun, S., Slickers, P., and Ehrlich, R. (2014). Rapid microarray-based DNA genosotyping of *Escherichia coli*. *Microbiol. Immunol.* 58, 77–86. doi: 10.1111/1348-0421.12120
- Greenfield, L. K., and Whitfield, C. (2012). Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr. Res.* 356, 12–24. doi: 10.1016/j.carres.2012.02.027
- Griffing, S. M., MacCannell, D. R., Schmidtke, A. J., Freeman, M. M., Hyytiä-Trees, E., Gerner-Smidt, P., et al. (2015). Canonical single nucleotide polymorphisms (SNPs) for high-resolution subtyping of Shiga-toxin producing *Escherichia coli* (STEC) O157: H7. *PLoS ONE* 10:e0131967. doi: 10.1371/journal.pone.0131967
- Hahm, B.-K., Maldonado, Y., Schreiber, E., Bhunia, A. K., and Nakatsu, C. H. (2003). Subtyping of foodborne and environmental isolates of *Escherichia coli* by multiplex-PCR, rep-PCR, PFGE, ribotyping and AFLP. *J. Microbiol. Methods* 53, 387–399. doi: 10.1016/S0167-7012(02)00259-2
- Hegde, N. V., Praul, C., Gehring, A., Fratamico, P., and DebRoy, C. (2013). Rapid O serogroup identification of the six clinically relevant Shiga toxin-producing *Escherichia coli* by antibody microarray. *J. Microbiol. Methods* 93, 273–276. doi: 10.1016/j.mimet.2013.03.024
- Hobbs, M., and Reeves, P. R. (1994). The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Mol. Microbiol.* 12, 855–856. doi: 10.1111/j.1365-2958.1994.tb01071.x
- Hyytiä-Trees, E., Smole, S. C., Fields, P. A., Swaminathan, B., and Ribot, E. M. (2006). Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). *Foodborne Pathog. Dis.* 3, 118–131. doi: 10.1089/fpd.2006.3.118
- Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., et al. (2015a). A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res.* 22, 101–107. doi: 10.1093/dnares/dsu043
- Iguchi, A., Iyoda, S., Seto, K., Morita-Ishihara, T., Sheutz, F., Ohnishi, M., et al. (2015b). *Escherichia coli* O-genotyping PCR: a comprehensive and practical platform for molecular O serogrouping. *J. Clin. Microbiol.* 53, 2427–2432.
- Ison, S. A., Delannoy, S., Bugarel, M., Nagaraja, T. G., Renter, D. G., Den Bakker, H. C., et al. (2016). Targeted amplicon sequencing for single-nucleotide-polymorphism genotyping of attaching and effacing *Escherichia coli* O26:H11 cattle strains via a high-throughput library preparation technique. *Appl. Environ. Microbiol.* 82, 640–649. doi: 10.1128/JCM.00321-15
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510.
- Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818
- Kaufmann, F. (1943). Ueber neue thermolabile Körperantigen der Colibakterien. *Acta Pathol. Microbiol. Scand.* 20, 21–44.
- Kaufmann, F. (1944). Zur serologie der coli-gruppe. *Acta Pathol. Microbiol. Scand.* 21, 20–45. doi: 10.1111/j.1699-0463.1944.tb00031.x
- Kaufmann, F. (1947). The serology of *E. coli* group. *J. Immunol.* 57, 71–100.
- Kotewicz, M. L., Jackson, S. A., LeClerc, J. E., and Cebula, T. A. (2007). Optical maps distinguish individual strains of *Escherichia coli* O157:H7. *Microbiology* 153, 1720–1733. doi: 10.1099/mic.0.2006/004507-0
- Kotewicz, M. L., Mammel, M. K., LeClerc, J. E., and Cebula, T. A. (2008). Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* 154, 3518–3528. doi: 10.1099/mic.0.2008/019026-0
- Krause, U., Thomson-Carter, F. M., and Pennington, T. H. (1996). Molecular epidemiology of *Escherichia coli* O157:H7 by pulsed-field gel electrophoresis and comparison with that by bacteriophage typing. *J. Clin. Microbiol.* 34, 959–961.
- Lacher, D. W., Gangiredla, J., Jackson, S. A., Elkins, C. A., and Feng, P. C. (2014). Novel microarray design for molecular serotyping of Shiga toxin-producing *Escherichia coli* strains isolated from fresh produce. *Appl. Environ. Microbiol.* 80, 4677–4682. doi: 10.1128/AEM.01049-14
- Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11
- Leung, K. T., Mackereth, R., Tien, Y.-C., and Topp, E. (2004). A comparison of AFLP and ERIC-PCR analyses for discriminating *Escherichia coli* from cattle, pig and human sources. *FEMS Microbiol. Ecol.* 47, 111–119. doi: 10.1016/S0168-6496(03)00254-X
- Lin, A., Nguyen, L., Lee, T., Clotilde, L. M., Kase, J. A., Son, I., et al. (2011). Rapid O serogroup identification of the ten most clinically relevant STECs by Luminex microbead-based suspension array. *J. Microbiol. Methods* 87, 105–110. doi: 10.1016/j.mimet.2011.07.019
- Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Reeves, P. R., et al. (2014). Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol. Rev.* 38, 56–89. doi: 10.1111/1574-6976.12034
- Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., et al. (2008). Structure and genetics of *Shigella* O antigens. *FEMS Microbiol. Rev.* 32, 627–653. doi: 10.1111/j.1574-6976.2008.00114.x
- Liu, Y., and Fratamico, P. (2006). *Escherichia coli* O antigen typing using DNA microarrays. *Mol. Cell Probes* 20, 239–244. doi: 10.1016/j.mcp.2006.01.001
- Liu, Y., Fratamico, P., DebRoy, C., Bumbaugh, A. C., and Allen, J. W. (2008). DNA sequencing and identification of serogroup-specific genes in the *Escherichia coli* O118 O antigen gene cluster and demonstration of antigenic diversity but only minor variation in DNA sequence of the O antigen clusters of *E. coli* O118 and O151. *Foodborne Pathog. Dis.* 5, 449–457. doi: 10.1089/fpd.2008.0096
- Liu, Y., Yan, X., DebRoy, C., Fratamico, P. M., Needleman, D. S., Li, R. W., et al. (2015). *Escherichia coli* O-antigen gene clusters of serogroups O62, O68, O131, O140, O142, and O163: DNA sequences and similarity between O62 and O68, and PCR-based serogrouping. *Biosensors (Basel)* 5, 51–68. doi: 10.3390/bios5010051
- Machado, J., Grimont, F., and Grimont, P. A. D. (2000). Identification of *Escherichia coli* flagellar types by restriction of the amplified fliC gene. *Res. Microbiol.* 151, 535–546. doi: 10.1016/S0923-2508(00)00223-0
- Manges, A. R., Harel, J., Masson, L., Edens, T. J., Portt, A., Reid-Smith, R. J., et al. (2015). Multilocus sequence typing and virulence gene profiles associated with *Escherichia coli* from human and animal sources. *Foodborne Pathog. Dis.* 12, 302–310. doi: 10.1089/fpd.2014.1860
- Manges, A. R., and Johnson, J. R. (2012). Food-borne origins of *Escherichia coli* causing extraintestinal infections. *Clin. Infect. Dis.* 55, 712–718. doi: 10.1093/cid/cis502
- Martin, I. E., Tyler, S. D., Tyler, K. D., Khakhria, R., and Johnson, W. M. (1996). Evaluation of ribotyping as epidemiologic tool for typing *Escherichia coli* serogroup O157 isolates. *J. Clin. Microbiol.* 34, 720–723.
- Miller, J. M. (2013). Whole-genome mapping: a new paradigm in strain-typing technology. *J. Clin. Microbiol.* 51, 1066–1070. doi: 10.1128/JCM.00093-13
- Mitchell, N. M., Johnson, J. R., Johnston, B., Curtis, R. III, and Mellata, M. (2015). Zoonotic potential of *Escherichia coli* isolates from retail chicken meat products and eggs. *Appl. Environ. Microbiol.* 81, 1177–1187. doi: 10.1128/AEM.03524-14
- Moreno, A. C. R., Guth, B. E. C., and Martinez, M. B. (2006). Can the fliC PCR-restriction fragment length polymorphism technique replace classic serotyping methods for characterizing the H antigen of enterotoxigenic *Escherichia coli* strains? *J. Clin. Microbiol.* 44, 1453–1458. doi: 10.1128/JCM.44.4.1453-1458.2006
- Namba, K., Yamashita, I., and Vonderviszt, F. (1989). Structure of the core and central channel of bacterial flagella. *Nature* 342, 648–654. doi: 10.1038/342648a0
- Nandanwar, N., Janssen, T., Kühl, M., Ahmed, N., Ewers, C., and Wieler, L. H. (2014). Extraintestinal pathogenic *Escherichia coli* (ExPEC) of human and avian origin belonging to sequence type complex 95 (STC95) portray indistinguishable virulence features. *Int. J. Med. Microbiol.* 304, 835–842. doi: 10.1016/j.ijmm.2014.06.009
- Nordstrom, L., Liu, C. M., and Price, L. B. (2013). Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. *Front. Microbiol.* 4:29. doi: 10.3389/fmicb.2013.00029
- Norman, K. N., Clawson, M. L., Strockbine, N. A., Mandrell, R. E., Johnson, R., Ziebell, K., et al. (2015). Comparison of whole genome sequences from human and non-human *Escherichia coli* O26 strains. *Front. Cell Infect. Microbiol.* 5:21. doi: 10.3389/fcimb.2015.00021
- Norman, K. N., Strockbine, N. A., and Bono, J. L. (2012). Association of nucleotide polymorphisms within the O-antigen gene cluster of *Escherichia coli* O26, O45,

- O103, O111, O121, and O145 with serogroups and genetic subtypes. *Appl. Environ. Microbiol.* 78, 6689–6703. doi: 10.1128/AEM.01259-12
- Ochman, H., and Selander, R. K. (1984). Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* 157, 690–693.
- Ørskov F., and Ørskov I. (1984). “Serotyping of *Escherichia coli*,” in *Methods in Microbiology*, Vol. 14, ed. T. Bergan (London: Academic Press), 43–112.
- Ørskov, F. I., Ørskov, F., Bettelheim, K. A., and Chandler, M. E. (1975). Two new *Escherichia coli* O antigens, O162 and O163, and one new H antigen, H56. Withdrawal of H antigen H50. *Acta Pathol. Microbiol. Scand.* 83, 121–124.
- Ørskov, I., Ørskov, F., Jann, B., and Jann, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol. Rev.* 41, 667–710.
- Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., et al. (2015). Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform. Biol. Insights* 9, 75–88. doi: 10.4137/BBI.S12462
- Ribot, E. M., Fair, M. A., Gautom, R., Cameron, D. N., Hunter, S. B., Swaminathan, B., et al. (2006). Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog. Dis.* 3, 59–67. doi: 10.1089/fpd.2006.3.59
- Russo, T. A., and Johnson, J. R. (2003). Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes Infect.* 5, 449–456. doi: 10.1016/S1286-4579(03)00049-2
- Samuel, G., and Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* 338, 2503–2519. doi: 10.1016/j.carres.2003.07.009
- Sarkar, S., Ulett, G. C., Totsika, M., Phan, M.-D., and Schembri, M. A. (2014). Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE* 9:e94786. doi: 10.1371/journal.pone.0094786
- Scheutz, F., Cheasty, T., Woodward, D., and Smith, H. R. (2004). Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new *E. coli* O groups that include verocytotoxin-producing *E. coli* (VTEC): O176, O177, O178, O179, O180 and O181. *APMIS*. 2004, 569–584.
- Selander, R. K., Caugant, K. A., Ochman, H., Musser, J. M., Gilmour, M. N., and Whittam, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* 51, 873–884.
- Shariat, N., and Dudley, E. G. (2014). CRISPRs: molecular signatures used for pathogen subtyping. *Appl. Environ. Microbiol.* 80, 430–439. doi: 10.1128/AEM.02790-13
- Singer, R. S. (2015). Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. *Front. Microbiol.* 6:28. doi: 10.3389/fmicb.2015.00028
- Smith, J. L., Fratamico, P. M., and Gunther, N. (2007). Extraintestinal pathogenic *Escherichia coli* (ExPEC). *Foodborne Pathog. Dis.* 4, 134–163. doi: 10.1089/fpd.2007.0087
- Stromberg, Z. R., Baumann, N. W., Lewis, G. L., Severt, N. J., Cernicchiaro, N., Renter, D. G., et al. (2015). Prevalence of enterohemorrhagic *Escherichia coli* O26, O45, O103, O111, O121, O145, and O157 on hides and preintervention carcass surfaces of feedlot cattle at harvest. *Foodborne Pathog. Dis.* 12, 631–638. doi: 10.1089/fpd.2015.1945
- Tseng, M., Fratamico, P., Bagi, L., Delannoy, S., Fach, P., Manning, S., et al. (2014). Diverse virulence gene content of Shiga toxin-producing *Escherichia coli* from finishing swine. *Appl. Environ. Microbiol.* 80, 6395–6402. doi: 10.1128/AEM.01761-14
- Vincent, C., Boerlin, P., Daignault, D., Dozois, C. M., Dutil, L., Galanakis, C., et al. (2010). Food reservoir for *Escherichia coli* causing urinary tract infections. *Emerg. Infect. Dis.* 16, 88–95. doi: 10.3201/eid1601.091118
- Wang, L., Rothmund, D., Curd, H., and Reeves, P. R. (2003). Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. *J. Bacteriol.* 185, 2936–2943. doi: 10.1128/JB.185.9.2936-2943.2003
- Wang, W., Perepelov, A. V., Feng, L., Shevelev, S. D., Wang, Q., Senchenkova, S. N., et al. (2007). A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology* 153, 2159–2167.
- Zhang, W., Qi, W., Albert, T. J., Motiwala, A. S., Alland, D., Hyttia-Trees, E. K., et al. (2006). Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. *Genome Res.* 16, 757–767. doi: 10.1101/gr.4759706

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

Copyright © 2016 Fratamico, DebRoy, Liu, Needleman, Baranzoni and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Analysis and Detection of *fliC_{H1}* and *fliC_{H12}* Genes Coding for Serologically Closely Related Flagellar Antigens in Human and Animal Pathogenic *Escherichia coli*

Lothar Beutin¹, Sabine Delannoy^{2*} and Patrick Fach²

¹ Department of Biology, Chemistry, Pharmacy, Institute for Biology - Microbiology, Freie Universität Berlin, Berlin, Germany,

² Université Paris-Est, Anses, Food Safety Laboratory, IdentityPath, Maisons-Alfort, France

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Sang Jun Lee,
Korea Research Institute of
Bioscience and Biotechnology, Korea
Timothy Read,
Emory University, USA

*Correspondence:

Sabine Delannoy
sabine.delannoy@anses.fr

Specialty section:

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

Received: 26 November 2015

Accepted: 25 January 2016

Published: 15 February 2016

Citation:

Beutin L, Delannoy S and Fach P
(2016) Genetic Analysis and Detection
of *fliC_{H1}* and *fliC_{H12}* Genes Coding
for Serologically Closely Related
Flagellar Antigens in Human and
Animal Pathogenic *Escherichia coli*.
Front. Microbiol. 7:135.
doi: 10.3389/fmicb.2016.00135

The *E. coli* flagellar types H1 and H12 show a high serological cross-reactivity and molecular serotyping appears an advantageous method to establish a clear discrimination between these flagellar types. Analysis of *fliC_{H1}* and *fliC_{H12}* gene sequences showed that they were 97.5% identical at the nucleotide level. Because of this high degree of homology we developed a two-step real-time PCR detection procedure for reliable discrimination of H1 and H12 flagellar types in *E. coli*. In the first step, a real-time PCR assay for common detection of both *fliC_{H1}* and *fliC_{H12}* genes is used, followed in a second step by real-time PCR assays for specific detection of *fliC_{H1}* and *fliC_{H12}*, respectively. The real-time PCR for common detection of *fliC_{H1}* and *fliC_{H12}* demonstrated 100% sensitivity and specificity as it reacted with all tested *E. coli* H1 and H12 strains and not with any of the reference strains encoding all the other 51 flagellar antigens. The *fliC_{H1}* and *fliC_{H12}* gene specific assays detected all *E. coli* H1 and all *E. coli* H12 strains, respectively (100% sensitivity). However, both assays showed cross-reactions with some flagellar type reference strains different from H1 and H12. The real-time PCR assays developed in this study can be used in combination for the detection and identification of *E. coli* H1 and H12 strains isolated from different sources.

Keywords: *E. coli*, molecular serotyping, *fliC* type H1 gene, *fliC* type H12 gene, STEC, ExPEC

INTRODUCTION

Strains belonging to the species of *Escherichia coli* are ubiquitous as commensals in the gut of humans and warm-blooded animals. Apart from their role as beneficial microbes, some *E. coli* strains are known to behave as human and animal pathogens, causing a wide spectrum of extraintestinal and enteric diseases, with urinary tract infection and diarrhea as most frequent (Kaper et al., 2004; Stenutz et al., 2006). Pathogenic and apathogenic *E. coli* cannot be discerned from each other by their morphology, cultural properties or fermentation reactions. As a consequence, serotyping is used since the 1940s as a diagnostic tool for identification of animal and human pathogenic *E. coli* strains (Orskov and Orskov, 1984).

E. coli serogroups are commonly defined by the antigenic properties of the lipopolysaccharide which is part of the outer membrane (O-antigen) (Stenutz et al., 2006). Motile *E. coli* strains can be

additionally typed for their flagellar filaments (H-antigen) (Orskov and Orskov, 1984). *E. coli* O- and H-antisera are usually produced by immunization of rabbits with respective reference strains (Orskov and Orskov, 1984; Edwards and Ewing, 1986). At present, 182 O-antigens and 53 H-antigens have been described (Scheutz et al., 2004; Scheutz and Strockbine, 2005). The resulting O:H serotype (for example O157:H7) is commonly used for describing *E. coli* isolates (Bettelheim, 1978; Orskov and Orskov, 1984).

Complete serotyping of *E. coli* is laborious and time-consuming and performed only in a few specialized reference laboratories worldwide. Moreover, cross-reactivity which is observed between some *E. coli* O-groups and H-types can complicate the interpretation of serotyping results. Last but not least, serotyping fails if autoagglutinating (O-antigen or H-antigen rough) and non-motile (NM) *E. coli* strains have to be examined (Orskov and Orskov, 1984; Edwards and Ewing, 1986). For these reasons, attempts were made to substitute serotyping by molecular typing of O-antigen and H-antigen encoding genes.

In the recent years, the nucleotide sequences of all known O and H-antigen genes in *E. coli* have been elucidated (Wang et al., 2003; Iguchi et al., 2015a). Molecular methods such as PCR and nucleotide sequencing have been successfully employed for typing of O- and H-antigen genes in *E. coli* (Beutin and Fach, 2014; Joensen et al., 2015; Iguchi et al., 2015b). Molecular serotyping was shown to be specific and sensitive and can substitute conventional serological detection of *E. coli* surface antigens (Bugarel et al., 2010; Fratamico et al., 2011; Clotilde et al., 2015; Iguchi et al., 2015b; Joensen et al., 2015). In contrast to serotyping, molecular detection of O- and H-antigen genes is easier and faster to perform and O-rough and non-motile strains can be typed on the basis of their O- and H-antigen genes (Beutin and Fach, 2014; Joensen et al., 2015).

We have previously investigated the genetic variability of flagellar types H19, H25 and H28 in *E. coli* (Beutin et al., 2015a,b). These flagellar types are widespread in strains belonging to numerous O-serogroups but are also associated with enterohemorrhagic *E. coli* O145:H25, O145:H28, and O121:H19 strains. By nucleotide sequence analysis of *fliC* (flagellin) genes encoding H19, H25, and H28 flagella we have observed a high genetic variability among *fliC*_{H19}, *fliC*_{H25}, and *fliC*_{H28} alleles, respectively. To some part, this sequence alterations were associated with some O-groups of strains which allowed the development of real-time PCR protocols for specific typing of flagellar variants encoded by enterohemorrhagic *E. coli* O145:H25, O145:H28, and O121:H19 strains (Beutin et al., 2015a,b). Such real-time PCR protocols were found useful for improvement of horizontal real-time PCR detection methods for EHEC from food samples (Beutin et al., 2015a,b).

In this work, we compared *E. coli fliC* genes that encode flagellar types H1 and H12. These flagellar types show a high serological cross-reactivity and cross-absorbed H1 and H12 antisera are used for definite H-typing (Orskov and Orskov, 1984; Edwards and Ewing, 1986). Moreover, three subtypes of H1 were detected by serological typing using factor specific antisera (Ratiner et al., 1995). Serological cross reactions may cause confounding results in diagnostic laboratories where absorbed

antisera are not available. The development of molecular typing procedures for reliable detection of H1 and H12 flagellar types could overcome this specific problem.

A clear discrimination between *E. coli* flagellar types H1 and H12 has a value for clinical diagnostics and for epidemiological investigations. Some human isolates of Shiga Toxin-producing *E. coli* (STEC) express H1 or H12 flagella (Scheutz and Strockbine, 2005). Moreover, flagellar type H1 is clinically significant as it is associated with worldwide occurring extraintestinal pathogenic *E. coli* (ExPEC) strains carrying capsular polysaccharides (O2:K2:H1, O4:K12:H1, O6:K2:H1, O6:K5:H1, O7:K1:H1, O15:K52:H1) that cause cystitis, pyelonephritis and urosepsis (Orskov and Orskov, 1985; Johnson et al., 1994, 2005, 2006; Olesen et al., 2009). Adherent-invasive *E. coli* (AIEC) O83:H1 strains were associated with Crohn's disease in human patients (Allen et al., 2008; Nash et al., 2010) and flagellar type H1 is associated with biofilm formation and invasive properties of AIEC strains (Eaves-Pyles et al., 2008; Martinez-Medina et al., 2009) as well as with intestinal colonization (Martinez-Medina and Garcia-Gil, 2014). Moreover, H1-type flagellum is a characteristic trait of Shiga toxin 2e-producing *E. coli* O139:H1 strains which are a major cause of edema disease in pigs (Tschape et al., 1992; Frydendahl, 2002; Fairbrother et al., 2005; Beutin et al., 2008). Conversely, the flagellar type H12 has not been associated with pathogenic *E. coli*, except from human enterotoxigenic O78:H12 and O128:H12 strains (Orskov and Orskov, 1977; Echeverria et al., 1982; Shaheen et al., 2004).

In this work we have analyzed the nucleotide sequences of *E. coli* H1 and H12 strains in order to detect characteristic *fliC* sequence alterations corresponding with these closely related H-types. Subsequently, we have developed a real-time PCR procedure for reliable discrimination of H1 and H12 flagellar types in *E. coli*. The protocol should be useful for diagnostic and epidemiological investigations of human and animal pathogenic strains of *E. coli*.

MATERIALS AND METHODS

Bacteria

E. coli strains used in this study were derived from the collections of the National Reference Laboratory for *E. coli* (NRL *E. coli*) at the Federal Institute for Risk Assessment (BfR) in Berlin, Germany and from the French Agency for Food, Environmental and Occupational Health and Safety (Anses) in Maisons-Alfort, France. *E. coli* strains used for specificity study included in particular the *E. coli* reference strains belonging to serogroups O1-O181 and H-types H1-H56 (Orskov and Orskov, 1984; Edwards and Ewing, 1986). All strains have been previously described for their serotypes and for virulence genes associated with STEC (Beutin et al., 2015a,b). All strains were grown overnight at 37°C in Luria broth, and DNA was extracted according to manufacturers instructions using InstaGene matrix (BioRad laboratories, Marnes-La-Coquette, France).

Real-time PCR assays were performed with an ABI 7500 instrument (Applied Biosystems, Foster City, CA, USA) in 25-µl reaction volumes, a LightCycler Nano (Roche Diagnostics,

Meylan, France) in 10 µl reaction volumes or with a LightCycler 1536 (Roche Diagnostics, Meylan, France) in 1.5-µl reaction volumes according to the recommendations of the suppliers. Primers and TaqMan probes were used at 300 nM final concentrations. The following thermal profile was applied to all instruments: enzyme activation at 95°C for 1–10 min as recommended followed by 40 cycles of denaturation at 95°C and annealing at 60°C.

PCR Detection and Mapping of *E. coli* O-Antigen and H-Antigen Genes

Mapping of *fliC* gene variants to their respective H-types was performed as previously described (Beutin et al., 2015a,b). Nucleotide sequence data obtained from thirteen *fliC*_{H1} and eight *fliC*_{H12} genes were used for designing TaqMan® real-time PCR probes and XS probes (minor groove binder replacement, Biolegio, Nijmegen, The Netherlands) and primers for specific detection of all genetic variants of thirteen *fliC*_{H1} and eight *fliC*_{H12} genes (this work). Real-time PCR probes and primers used in this work were designed with the software Primer Express V3.0 (Applied Biosystems) and are described in Table 1.

Nucleotide Sequencing

The nucleotide sequence of the PCR products were determined as described (Beutin et al., 2015b) and analyzed with the Accelrys DS Gene software package (Accelrys Inc., USA). The nucleotide sequences of the respective products for *fliC* homologs were determined and have been submitted to European Nucleotide Archive (ENA). The GenBank Accession numbers are listed in Table 2.

RESULTS

Sources and Properties of *E. coli* H1 and H12 Strains

The *E. coli* H1 and H12 strains investigated in this study were from human, animal, food, and environmental sources (Table 3). The thirty-one flagellar type H1 strains were associated with 10 different *E. coli* O-serogroups, O-rough and O-untypable strains and originated from healthy and diseased humans and animals and from food. The thirty-eight H12 strains divided into thirteen different O-groups of *E. coli*, and in O-untypable and O-rough strains. The H12 strains were from healthy and diseased humans and animals, from food and the environment. Production of Shiga-toxins (Stx) was found in 16 (42.1%) of the H12 strains and associated with five different O-groups. Fourteen (45.2%) of the *E. coli* H1 strains produced Stx, however most of these were from pigs with edema disease (O139:H1, Or:H1) and harbored the *stx2e* gene. O:H types known to be associated with *E. coli* causing extraintestinal infections of humans (O2:H1, O4:H1, O6:H1, O25:H1) were detected among the investigated H1 strains. Interestingly, strains belonging to these serotypes originated not only from humans but also from animals and food. Certain strains belonged to serotypes which have not been previously associated with clinical disease and their role of pathogens for humans and animals is not yet known.

TABLE 1 | Primers and probes for real-time PCR detection of *E. coli* flagellar types H1 and H12.

Target gene	Forward primer, reverse primer and probe sequences (5'–3') ^a	Length and location within 21 sequences listed in Table 2 (5'–3')
<i>fliC</i> _{H1}	AGGACGAAATCAAATCCCGTCT ACGGTTCGATGAAAATTCAGGTT [6FAM]-GACCGCGTATCC GGTCA-[BHQ1] ^a	338–359 ^b 422–444 ^c 370–386 ^d
<i>fliC</i> _{H12}	TCCATTGAGGACGAAATCAAATC CGTGAACGTACTGGCGAAAG [6FAM]-GTATCTGGCQAG ACCCA-[BHQ1] ^a	331–353 ^b 402–421 ^e 376–392 ^f
<i>fliC</i> _{H1/H12}	TGATGGTGAAATGACTACAATTGGT GGTAACTGTTGATTCTGGAAGTGGT [6FAM]-CGAAGTATTCAATCG ATGCTAACAACGGCA-[BHQ1]	1329–1353 ^g 1395–1419 ^g 1363–1393 ^g

^aXS probes (MGB-replacement) were used for *fliC*_{H1} and *fliC*_{H12} specific real-time PCRs.

^bForward primer conserved in all analyzed *fliC*_{H1} and *fliC*_{H12} sequences.

^c*fliC*_{H1} reverse primer: one mismatch at position 429: *fliC*_{H1} = G, *fliC*_{H12} = A (underlined).

^d*fliC*_{H1} probe: one mismatch at position 381: *fliC*_{H1} = C, *fliC*_{H12} = T; and position 384: *fliC*_{H1} = T, *fliC*_{H12} = C (underlined).

^e*fliC*_{H12} reverse primer: one mismatch at position 402 with *fliC*_{H1} = T (4/13 strains), *fliC*_{H12} = C (underlined).

^f*fliC*_{H12} probe: one mismatch at position 381: *fliC*_{H12} = T, *fliC*_{H1} = C; and position 384: *fliC*_{H12} = C, *fliC*_{H1} = T (underlined).

^gConserved in all 21 *fliC*_{H1} and *fliC*_{H12} sequences from Table 2.

Nucleotide Analysis of *E. coli fliC*_{H1} and *fliC*_{H12} Genes

The nucleotide sequences of the reference strains (Orskov and Orskov, 1984) for *E. coli* flagellar antigens H1 (strain Su1242, GenBank accession AB028471.1) and H12 (Bi 316-42, GenBank accession AY249997) (Wang et al., 2003) have been published previously. The length of coding sequence of each *fliC*_{H1} and *fliC*_{H12} gene is 1788 nucleotides and both sequences have 97.5% identity (44 nucleotide exchanges) at the nucleotide level and 98.98% identity and 99.16% similarity at the amino acid level (7 amino acids (aa) exchanges). Additional *fliC* nucleotide sequences from six *E. coli* H1 and five *E. coli* H12 strains were obtained in this work (Table 2). These sequences were compared with seven *fliC*_{H1} sequences and three *fliC*_{H12} sequences already available in GenBank (Table 2). All 21 H1 or H12 flagellin genes have a 1788 nucleotides length that codes for 595 amino acid residues.

A cluster analysis performed with thirteen *fliC*_{H1} and eight *fliC*_{H12} sequences is shown in Figure 1. Four different genotypes were detected among the thirteen *fliC*_{H1} strains. Uropathogenic *E. coli* O2:H1, O6:H1, O25:H1, and AIEC O83:H1 strains were identical for their *fliC*_{H1} sequences and assigned to a large cluster composed by eight strains. A smaller cluster was formed by five *fliC*_{H1} strains; four of these were Stx2e producing O139:H1 causing edema disease in pigs.

Six different genotypes were found among the eight *fliC*_{H12} strains. Identical *fliC*_{H12} sequences were only found between two O9:H12 strains and each one O55:H12 and O153:H12 strain, respectively.

TABLE 2 | *Escherichia coli* strains used for nucleotide sequencing of *fliC*_{H1} and *fliC*_{H12} genes and corresponding sequences obtained from GenBank.

Strain	Serotype	<i>fliC</i> gene GenBank accession no	Pathotype	Source and References
CB11962	O20:H12	LN877748 ^e	STEC	Calves feces, Germany, 2009, this work
CB13385	O9:K9:H12	LN877749 ^e	No data	Chicken meat, Germany 2011, this work
Bi316-42	O9:K9:H12	AY249997	ExPEC	Orskov and Orskov, 1984; Wang et al., 2003
NX9861	O157:H12	AY337474	No data	China, 2003, unpublished
90	O157:H12	AY337471	No data	China, 2003, unpublished
CB11070	O1:H12	LN877750 ^e	No data	Pig feces, Germany, 2007, this work
CB12026(07QMA185.1)	O153:H12	LN877751 ^e	STEC	Beef, France, 2009, this work
CB12530	O55:H12	LN877752 ^e	STEC	Martin and Beutin, 2011
Su1242	O2:K2:H1	AB028471	ExPEC	Orskov and Orskov, 1984
CFT073	O6:H1	AE014075	ExPEC	Welch et al., 2002
ABU83972	O25:H1	CP001671 ^a	ExPEC	Zdziarski et al., 2010
CB13658	O6:H1	LN877753 ^e	No data	Pig intestine, Germany, 2011, this work
ATCC25922	O6:H1	CP009072	Human, clinical	Minogue et al., 2014
NRG 857C	O83:H1	CP001855	AIEC	Allen et al., 2008
LF82	O83:H1	CU651637 ^b	AIEC	Martinez-Medina et al., 2009
Ec614	O157:H1	JF308285 ^c	Beef, no data	Goulter et al., 2010
CB13179	O15:H1	LN877754 ^e	Human, ESBL-producer ^d	Geser et al., 2012
CB295	O139:H1	LN877755 ^e	STEC	Hampson et al., 1988
CB13050(D3648)	O139:H1	LN877756 ^e	STEC	Scheutz et al., 2012
CB13107	O139:H1	LN877757 ^e	STEC	Switzerland, pig intestine, 2011, this work
CB15303	O139:H1	LN877758 ^e	STEC	France, edema disease pig, 2014, this work

^a The whole genome sequence of the *E. coli* strain ABU 83972 (GenBank: CP001671.1) is available (Zdziarski et al., 2010). The O-serogroup of this strain was not reported but its *wzx* gene (position 2372093–2373353) is >99% similar to *wzx* of *E. coli* O25 strains E47a (GenBank GU014554.1) (Wang et al., 2010). Therefore, we classified ABU 83972 here as an O25:H1 strain.

^b The whole genome sequence of *E. coli* strain LF82 is available (GenBank: CU651637.1). The O-serogroup of this strain is not reported but its *wzx* gene (position 2127428–212804) is identical to the *wzx* gene of *E. coli* O83:H31 strain H17a GenBank: KJ778808.1 (unpublished) and of *E. coli* O83:H1 strain NRG857c (GenBank: CP001855.1) (Allen et al., 2008). Therefore, we suggest that LF82 is an O83:H1 strain.

^c The *fliC* sequence deposited under GenBank JF308285 is derived from strain EC614 reported as O157:H1 (Goulter et al., 2010). By Blast search, it is 100% identical to the *fliC* sequence of the H1 reference strain Su1242 (GenBank accession AY249997). Therefore, the flagellar type of EC614 was classified as H1.

^d Multiresistant, extended-spectrum-lactamase (ESBL)-producing *E. coli* from healthy human carrier.

^e The *fliC* sequence was determined in this study.

Amino acid Alterations between Flagellar Antigens H1 and H12 in *E. coli* Strains

An alignment of the amino acid sequences of thirteen *fliC*_{H1} and eight *fliC*_{H12} strains is shown in Table S1. All translation products had a length of 595 amino acids (aa). The eight *fliC*_{H12} strains were showing only few alterations with one or more of the strains at aa positions 249, 258, 339, and 472 (99.2% similarity) (Table S1), generating six different protein sequences (Figure 2). The thirteen *fliC*_{H1} strains split into three protein sequences (Figure 2) differing at positions 258, 431, and 481 (99.5% similarity) (Table S1). The aa changes were all located in the variable region of *fliC* encoding flagellar antigen specificity (Wang et al., 2003). Differences in the aa sequence which could distinguish between all investigated *fliC*_{H1} and *fliC*_{H12} strains, respectively, were found at positions 302 (Glu/Lys), 340 (Asn/Lys), 361 (Gly/Asp), 391 (Thr/Lys), 396 (Asn/Asp), and 430 (Asn/Lys). The six flagellar type specific aa sequence differences

were all located in the variable region of the *fliC*_{H1} and *fliC*_{H12} genes.

Development and Evaluation of Real-Time PCR Assays for Identification of *E. coli fliC*_{H1} and *fliC*_{H12} Strains

The close similarity between *E. coli fliC*_{H1} and *fliC*_{H12} translation products explains the serological cross-reactivity which was previously described for H1 and H12 antigens (Orskov and Orskov, 1984; Edwards and Ewing, 1986). As specific differences were found that distinguish between *fliC*_{H1} and *fliC*_{H12} sequences, molecular detection of the respective *fliC* genes could be more suitable than serotyping for clear identification of H1 and H12 strains of *E. coli*.

Based on the sequence data obtained for *E. coli fliC*_{H1} and *fliC*_{H12} genes we developed a TaqMan real-time PCR assay for common detection of *fliC*_{H1} and *fliC*_{H12} genes as well as

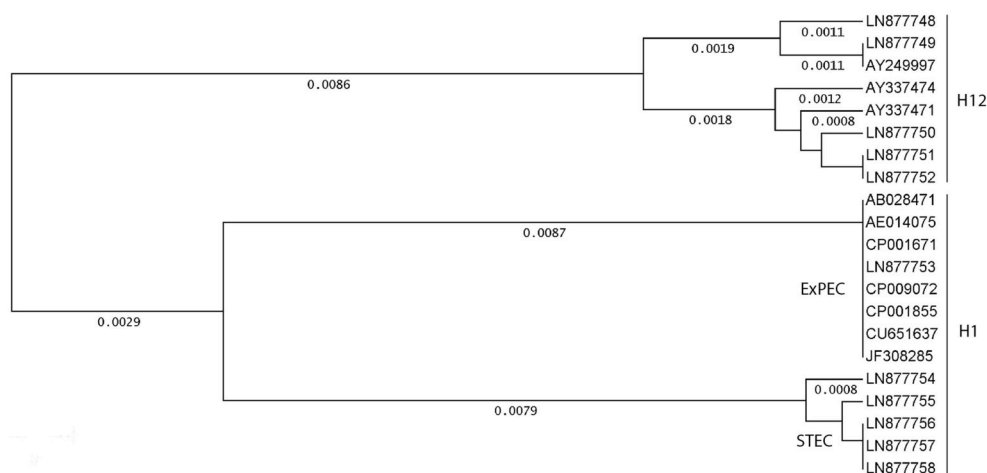


FIGURE 1 | Genetic relationships between *fliC*_{H1} and *fliC*_{H12} genes in different strains and serotypes of *E. coli*. Cluster analysis was performed using eight *fliC*_{H12} and thirteen *fliC*_{H1} genes listed in **Table 2**. GenBank accession numbers are indicated for orientation. The unweighted-pair group method using average linkages was used as a tree-building mode, and the distances were calculated according to Tajima and Nei (1984) using the Accelrys DS Gene software package.

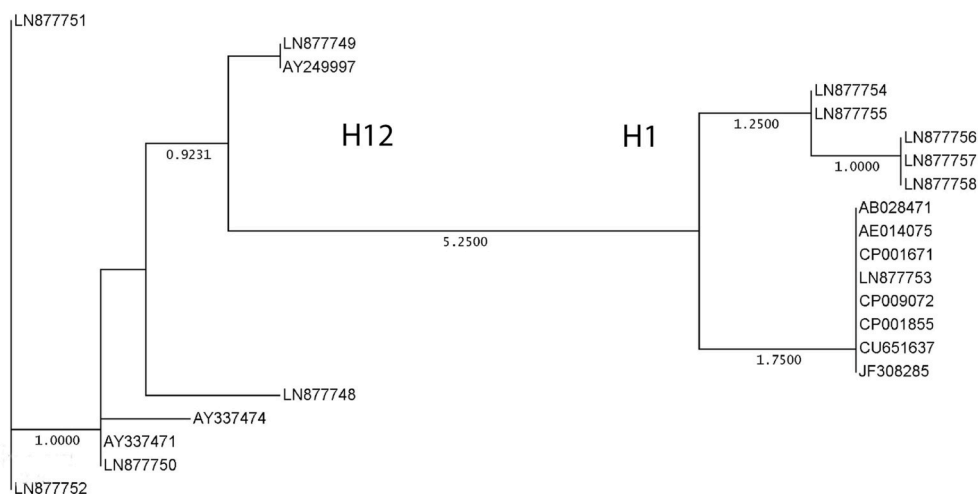


FIGURE 2 | Genetic relationships between translation products of thirteen *fliC*_{H1} and eight *fliC*_{H12} genes in different strains and serotypes of *E. coli* listed in **Table 2.** GenBank accession numbers are indicated for orientation. The neighbor joining method with absolute differences (best tree) was used as a tree-building mode (Nei, 1996) using the Accelrys DS Gene software package.

real-time PCR assays for specific detection of *fliC*_{H1} and *fliC*_{H12}, respectively (**Table 1**). Short-length XS-probes (minor groove binder replacement) had to be employed to develop real-time PCR assays specific for *fliC*_{H1} and *fliC*_{H12} sequences (**Table 1**). We used two nucleotide substitutions between the sequences of *fliC*_{H1} and *fliC*_{H12} to design specific probes (**Table 1**).

The assays were first tested for sensitivity and specificity on 31 *E. coli* H1 and 38 *E. coli* H12 strains (**Table 4**) as well as on the *E. coli* H-type reference strains (H1-H56) (Orskov and Orskov, 1984; Edwards and Ewing, 1986). The real-time PCR for common detection of *fliC*_{H1} and *fliC*_{H12} reacted with all tested *E. coli* H1 and H12 strains (**Table 4**) and not with any of the reference strains encoding all other flagellar antigens than H1 and H12.

The *fliC*_{H1} and *fliC*_{H12} gene specific assays detected all *E. coli* H1 and all *E. coli* H12 strains, respectively (**Table 4**). However, both assays showed cross-reactions with some flagellar type reference strains different from H1 and H12. With the *fliC*_{H1} real-time PCR, cross-reactions were observed with H6, H7, H15, H20, H34, H37, H41, H45, H46, H49, and H52 strains. The *fliC*_{H12} specific PCR reacted also with H7, H28, H31, and H41 strains (**Table 5**). Although the overall sequences of the *fliC* genes of H-types cross-reacting with the *fliC*_{H1} and *fliC*_{H12} real-time PCR are widely different from those of *fliC*_{H1} and *fliC*_{H12}, they show high local similarities with the primers and probes sequences. In cases of cross reactivity, no or only minor differences (0–3 mismatches) were found between target-sequences and *fliC*_{H1}

TABLE 3 | Source and origin of *E. coli* H1 and H12 strains.

Serotype ^a	Nos. of strains	Source	Origin/References
O2:H1	1	Calves feces, diarrhea	Germany, 2010
O2:K2:H1	1	Human blood	Orskov and Orskov, 1984
O4:H1	1	Raw milk cheese	Germany, 2010
O6:H1	4	Pig feces, diarrhea (1), Human (3)	Germany, 2011 Germany, 2009
O6:K5:H1	1	Human feces	Reister et al., 2014
O15:H1	3	Rabbit (2) Human (1)	Switzerland, 2007 Geser et al., 2012
O22:H1	2	Goat cheese (1) Human peritoneum (1)	Germany, 2013 Orskov and Orskov, 1984
O25:H1	1	Dog feces	Germany, 2011
O77:H1	1	salad	Germany, 2009
O79:H1	1	hare	Germany, 2007
O139:H1	11 ^b	Pig feces/organs, edema disease Wild boar feces/organs, edema disease	Orskov and Orskov, 1984; Beutin et al., 2008 France, 2013
O149:H1	2 ^c	Beef	Germany, 2011
ONT:H1	1	human	Germany, 2011
Or:H1	1 ^b	Pig feces, edema disease	Germany, 2014
O1:H12	1	Pig feces	Germany 2007
O9:K9:H12	5	Human peritoneum (1) Surface water (3) Chicken meat (1)	Orskov and Orskov, 1984 Germany, 2013 Germany, 2013
O9:H12	4	Pork	Martin and Beutin, 2011
O11:H12	2	Human (1) Pig feces (1)	Geser et al., 2012 Germany, 2009
O20:H12	2 ^d	Calves feces/organs, diarrhea	Germany, 2009
O49:H12	1	Human urine	Orskov and Orskov, 1984
O55:H12	2 ^e	Milk, beef	Martin and Beutin, 2011
O79:H12	1	Surface water	Germany, 2011
O98:H12	1	Pork	Germany, 2013
O104:H12	2	Human, diarrhea Surface water	Miko et al., 2013 Germany, 2013
O118:H12	3 ^d	Human, diarrhea	Pierard et al., 1998; Beutin et al., 2004
O136:H12	4 ^e	Milk Cattle feces	Martin and Beutin, 2011 France, 1998 Canada, 2012
O153:H12	3 ^e	beef	Martin and Beutin, 2011
O157:H12	3	Human Pig	Germany, 2007 Kaufmann et al., 2006
ONT:H12	2	Milk Surface water	Germany, 2014 Germany, 2013
Or:H12	2 ^e	Cattle feces	Germany, 2010

^a This list includes serotype reference strains: Nissle 1917 (O6:K5:H1) (Reister et al., 2014), EH250 (O118:H12), (Scheut et al., 2012), Su 1242 (O2:K2:H1), E14a (O22:H1), CDC 63-57 (O139:H1), Bi316-42 (O9:K9:H12), U12-41 (O49:H1) (Orskov and Orskov, 1984).

^b Positive for *stx2e*.

^c Positive for *stx1d*.

^d Positive for *stx2*.

^e Positive for *stx1*.

TABLE 4 | Detection of different *E. coli* H1 and H12 strains belonging to different O-serogroups by *fliC*_{H1}, *fliC*_{H12} and *fliC*_{H1/H12} Real Time PCR assays.

Serotype ^b	Nos. of strains	CT-values ^a <i>fliC</i> _{H1}	CT-values ^a <i>fliC</i> _{H12}	CT-values ^a <i>fliC</i> _{H1/H12}
O2:H1	1	21.1	–	22.2
O2:K2:H1 ^b	1	21.1–24.3	–	22.2–22.9
O4:H1	1	24.7	–	27.1
O6:H1	5	21.1–24.3	–	21.9–25.3
O15:H1	3	19.6–25.5	–	21.2–24.0
O22:H1	2	19.6–22.0	–	22.6–24.2
O25:H1	1	22.5	–	24.5
O77:H1	1	21.8	–	23.1
O79:H1	1	24.5	–	24.7
O139:H1	11	16.5–23.7	–	15.0–24.6
O149:H1	2	213–21.4	–	23.5–24.8
ONT:H1	1	22.6	–	22.7
Or:H1	1	20.1	–	21.9
O1:H12	1	–	20.9	24.2
O9:K9:H12 ^b	5	–	18.6–23.2	21.6–23.1
O9:H12	4	–	17.0–17.9	16.0–22.6
O11:H12	2	–	18.2–22.4	21.7–22.6
O20:H12	2	–	15.2–16.9	20.4–21.1
O49:H12	1	–	19.6	24.5
O55:H12	2	–	20.1–22.8	22.5–22.7
O79:H12	1	–	22.8	23.0
O98:H12	1	–	22.6	24.3
O104:H12	2	–	20.9–23.6	22.1–23.4
O118:H12	3	–	18.0–23.1	17.1–23.4
O136:H12	4	–	16.8–23.6	15.9–23.2
O153:H12	3	–	18.3–20.9	22.2–23.2
O157:H12	3	–	18.7–23.4	21.7–23.7
ONT:H12	2	–	21.7–22.8	22.5–23.4
Or:H12	2	–	19.2–20.7	23.0–23.3

^a Range of real time PCR cycle thresholds. Negative reactions are indicated with the “–” sign.

^b Reference strain Orskov non-motile and the *fliC*-genotype was detected by nucleotide sequencing of *fliC* PCR products.

and *fliC*_{H12}, primers and probes (Table 5). Three and more mismatches were found in cases of negative real-time PCR results.

In respect to these findings, the assays were then tested on a second panel of 78 strains comprising strains with H-types previously found to cross-react with *fliC*_{H1} or *fliC*_{H12} PCR assays as well as strains from O-groups that can be found associated with H1 and H12, but with H-types different from H1 and H12 (Table 6).

None of the 78 strains with H-types different from H1 and H12 reacted with the common *fliC*_{H1} / *fliC*_{H12} real-time PCR assay. Cross reactions with the *fliC*_{H1} real-time PCR-assay were observed with H6 (9/9), H49 (3/3), H31 (1/2), H34 (2/7), H41 (1/2), H45 (3/4) as well as with one O6:H4 strain. Weak cross-reactions were also observed with one O2:H25 strain and one O153:H25 strain. Cross-reactions with the *fliC*_{H12} real-time PCR-assay were observed with H7 (9/9), H28 (6/6), H31 (1/2),

TABLE 5 | Cross-reactions of *fliC*_{H1} and *fliC*_{H12} real time PCR assays with other flagellar types of *E. coli*.

Reference strain ^a	H-type	GenBank Accession No.	Detector tested ^b	CT-value ^c	Mismatch FP/P/RP ^d
A20	H6	AY249991.1	<i>fliC</i> _{H1}	26.8	0/0/0
			<i>fliC</i> _{H12}	–	0/2/1
U5-41	H7	AB028474.1	<i>fliC</i> _{H1}	26.0	0/1/0
			<i>fliC</i> _{H12}	24.5	0/1/1
E39a	H15	AY249999.1	<i>fliC</i> _{H1}	28.2	1/1/1
			<i>fliC</i> _{H12}	–	3/2/2
H3306	H20	AY250003.1	<i>fliC</i> _{H1}	27.0	0/0/1
			<i>fliC</i> _{H12}	–	2/2/2
HW30	H28	AY337469.1	<i>fliC</i> _{H1}	–	0/2/1
			<i>fliC</i> _{H12}	21.9	0/0/0
HW33	H31	AF345849.1	<i>fliC</i> _{H1}	–	0/2/4
			<i>fliC</i> _{H12}	22.9	0/0/0
BP 12665	H34	AY250016.1	<i>fliC</i> _{H1}	20.9	0/0/0
			<i>fliC</i> _{H12}	–	0/2/1
P11a	H37	AY250017.1	<i>fliC</i> _{H1}	26.4	1/0/1
			<i>fliC</i> _{H12}	–	3/2/1
RVC1787	H41	AY250020.1	<i>fliC</i> _{H1}	27.1	0/1/1
			<i>fliC</i> _{H12}	24.0	0/1/1
4106-54	H45	AY250023.1	<i>fliC</i> _{H1}	25.5	0/0/0
			<i>fliC</i> _{H12}	–	0/3/1
5306-56	H46	AY250024.1	<i>fliC</i> _{H1}	27.6	0/0/1
			<i>fliC</i> _{H12}	–	0/1/2
2147-59	H49	AY250026.1	<i>fliC</i> _{H1}	24.6	0/0/0
			<i>fliC</i> _{H12}	–	0/3/1
C2187-69	H52	AY250028.1	<i>fliC</i> _{H1}	26.8	0/1/1
			<i>fliC</i> _{H12}	–	0/3/1
Su1242	H1	AB028471.1	<i>fliC</i> _{H1}	21.1–24.3	0/0/0
			<i>fliC</i> _{H12}	–	0/2/1
Bi316/42	H12	AY249997.1	<i>fliC</i> _{H1}	–	0/2/1
			<i>fliC</i> _{H12}	18.6–23.6	0/0/0

^aH-type reference strains (Orskov and Orskov, 1984).^bAs listed in **Table 1**.^cMean of real-time cycle threshold (CT-values) calculated from duplicate PCRs. Negative reactions are indicated with the “–” sign.^dNumber of mismatches found between real-time detector sequence and target gene sequence. FP, forward primer; P, gene probe; RP, reverse primer.

H34 (2/7), H41 (1/2), as well as one O20:H9, one O55:H19 and one O153:H14 strains. In contrast to the respective reference strains, cross-reactions were not observed with either real-time PCR-assay with two other H15 and one H52 strain tested (**Tables 5, 6**). We do not know if these three strains show further differences in the PCR-target region which could explain these findings.

Overall, molecular typing of *E. coli* H1 and H12 strains requires first identification of H1/H12 strains with the common *fliC*_{H1}/*fliC*_{H12} real-time PCR assay, followed by specific identification of *fliC*_{H1} and *fliC*_{H12}, by their respective real-time PCR-assays. The real-time PCR for common detection of *fliC*_{H1} and *fliC*_{H12} was found 100% sensitive and 100% specific. The *fliC*_{H1} and *fliC*_{H12} gene specific assays were found 100% sensitive as they detected all *E. coli* H1 and all *E. coli* H12 strains,

respectively. When used exclusively on H1 and H12 strains (as identified by the common primers/probe set in a first step), the *fliC*_{H1} and *fliC*_{H12} gene specific assays were found 100% specific. Thus, 100% of H1 and H12 strains would be accurately typed with this system.

DISCUSSION

The genetically and serologically closely related flagellar antigens H1 and H12 were found in heterogeneous types of *E. coli* strains belonging to 26 different O-serogroups, O-untypable and O-rough strains. With one exception (O79:H1 and O79:H12), H1 and H12 strains did not share common O-serogroups which would indicate that flagellar types H1 and H12 have separated from each other not very recently in evolution. They may have evolved independently following rearrangements in the O-group loci of ancestor strains carrying the closely related H1/H12 flagellar types and do not directly derive from a common O-group ancestor.

By comparing nucleotide sequences of *fliC* genes from thirteen H1 and eight H12 strains we identified six H-type specific aa changes at positions 302, 340, 361, 391, 396, and 430. All these are located in the variable part of flagellin determining antigen specificity (Wang et al., 2003). As these changes are characteristic for the respective flagellar antigen, we suppose them to determine the antigen specificities of H1 and H12. The few other aa changes detected in some H1 and H12 strains might thus not be significant as specific characteristics of H1 or H12 types. However, such aa-changes could explain the finding of serological subtypes of H1 which were detected using factor specific H-antisera (Ratiner et al., 1995).

Interestingly, the genetic distance between *fliC*_{H1} (Su1242, GenBank accession AB028471.1) and *fliC*_{H12} sequences (Bi 316-42, GenBank accession AY249997) (97.5% similarity) is less than that found between different alleles of *fliC*_{H28} (92.0% similarity) (Beutin et al., 2015b). It is slightly bigger than the distance found among different alleles of *fliC*_{H19} (98.5% similarity) (Beutin et al., 2015a). Multiple allelic types of *fliC* were also detected in *E. coli* H6, H7, H8, H25, and H40 strains, respectively (Reid et al., 1999; Wang et al., 2000; Beutin and Strauch, 2007; Beutin et al., 2015b). Already, a considerable number of serological cross-reactions were observed when flagellar types H1–H56 were compared (Orskov and Orskov, 1984; Edwards and Ewing, 1986). Some of these flagellar antigens (H1/H12, H8/H40, H11/H21, and H37/H41) are so closely related that the use of cross-absorbed antisera is needed to obtain unambiguous serotyping results (Edwards and Ewing, 1986).

The presence of allelic subtypes within *fliC* genes encoding different H-types of *E. coli* and the finding that different flagellar types are serologically cross-reacting may complicate *E. coli* strain typing using H-antisera. The use of molecular typing procedures, such as real-time PCR can solve the typing problem caused by serologically closely related H-antigens, as we have shown for H1 and H12 in this work. Using primer express V3.0 software, it was not possible to design real-time PCRs specific exclusively for *fliC*_{H1} and *fliC*_{H12}, respectively. For this reason, we employed a two-step real-time detection procedure. The first step uses a real-time PCR highly specific for both H1 and H12

TABLE 6 | Reaction of the *fliC_{H1/H12}*, *fliC_{H1}* and *fliC_{H12}* real-time PCR assays with non-H1 and non-H12 strains.

Serotype	Number of strains	Ct-values ^a <i>fliC_{H1/H12}</i>	Ct-values ^a <i>fliC_{H1}</i>	Ct-value ^a <i>fliC_{H12}</i>
O33:H6	1	–	20.6	–
O40:H6	1	–	19.35	–
O55:H6	1	–	17.2	–
O63:H6	1	–	18.98	–
O113:H6	2	–	19.38–19.46	–
O125:H6	1	–	24.29	–
O126:H6	1	–	20.17	–
O127:H6	1	–	21.78	–
O41:H7	1	–	–	21.11
O55:H7	3	–	–	22.35–26.16
O153:H7	1	–	–	18.21
O157:H7	4	–	–	18.46–24.7
O23:H15	1	–	–	–
O157:H15	1	–	–	–
O28:H28	1	–	–	18.73
O91:H28	1	–	–	16.34
O110:H28	1	–	–	15.53
O116:H28	1	–	–	30.55
O145:H28	1	–	–	14.99
OX185:H28	1	–	–	17.18
O51:H49	1	–	21.1	–
O114:H49	1	–	20.65	–
O181:H49	1	–	20.6	–
O45:H31	1	–	19.44	–
O179:H31	1	–	–	19.56
O6:H34	2	–	–	–
O86:H34	1	–	21.4	–
O142:H34	1	–	21.99	–
O145:H34	1	–	–	21.67
O132:H34	1	–	–	–
O132:H34	1	–	–	20.84
O76:H41	1	–	–	20.75
O17/77:H41	1	–	21.93	–
O8:H45	1	–	–	–
O121:H45	1	–	22.32	–
O157:H45	1	–	24.34	–
O186:[H45]	1	–	20.92	–
O119:[H52]	1	–	–	–
O2:H8	1	–	–	–
O2:H25	1	–	27.15	–
O2:H27	1	–	–	–
O2:H40	1	–	–	–
O4:H5	1	–	–	–
O4:H16	1	–	–	–
O6	2	–	–	–
O6:H4	1	–	24.17	–
O6:H10	1	–	–	–
O7:H4	1	–	–	–
O15:H2	1	–	–	–

(Continued)

TABLE 6 | Continued

Serotype	Number of strains	Ct-values ^a <i>fliC_{H1/H12}</i>	Ct-values ^a <i>fliC_{H1}</i>	Ct-value ^a <i>fliC_{H12}</i>
O15:H11	1	–	–	–
O15:H16	1	–	–	–
O15:H21	1	–	–	–
O139:H4	1	–	–	–
O139:H19	1	–	–	–
O128:H2	1	–	–	–
O128:H8	1	–	–	–
O20:H9	1	–	–	23.23
O20:H30	1	–	–	–
O20:H33	1	–	–	–
O20:NM	1	–	–	–
O55:H19	1	–	–	23.43
O55:H21	1	–	–	–
O55:H51	1	–	–	–
O118:H2	1	–	–	–
O118:H5	1	–	–	–
O118:H8	1	–	–	–
O118:H16	1	–	–	–
O153:H14	1	–	–	20.86
O153:H21	1	–	–	–
O153:H25	1	–	28.37	–

^aRange of real time PCR cycle thresholds. Negative reactions are indicated with the “–” sign.

strains, followed by subtyping of H1/H12-positive strains with the respective *fliC_{H1}* and *fliC_{H12}* specific real-time PCRs. Short probe sequence lengths as obtained with minor groove binder (MGB) or MGB-replacements (XS-probe) are needed to ensure specificity between closely similar DNA-targets as previously shown for *fliC_{H19}* allelic discrimination (Beutin et al., 2015a). The PCRs could be used in parallel for examination of large number of isolates using high throughput PCR platforms as described previously for analysis of large numbers of *Clostridia* and *E. coli* strains (Delannoy et al., 2013; Woudstra et al., 2013).

Unambiguous typing of *fliC_{H1}* and *fliC_{H12}* sequences is of interest for clinical and epidemiological investigations since some H1 and H12 strains were shown to play a role as pathogens in humans and animals.

More than one third of investigated H1 and H12 strains produced Shiga toxins. Strains showing O:H types characteristic for ExPEC associated with human diseases (O2:H1, O4:H1, O6:H1, O15:H1) were detected in this work. Interestingly, these were not only from humans but also found in animals and food. It was previously described that animals, food and water can be a source of pandemic ExPEC strains (Jakobsen et al., 2010; Riley, 2014; Gomi et al., 2015; Singer, 2015). Flagellar type H12 strains encompass mainly STEC (O20:H12, O55:H12, O118:H12, O136:H12, O153:H12, and Or:H12) and were isolated from diseased animals and humans, food and the environment (Scheut and Strockbine, 2005).

The specific molecular detection of H1 and H12 flagellins as described in this study will be useful for diagnosis and for source attribution of human and animal pathogenic ExPEC and STEC strains in outbreaks and sporadic cases of infection.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: LB, SD, PF. Performed the experiments: LB, SD. Analyzed the data: LB, SD, PF. Contributed reagents/materials/analysis tools: LB, SD, PF. Wrote the paper: LB, SD, PF. Critical revision of the paper for important intellectual content: LB, SD, PF.

REFERENCES

- Allen, C. A., Niesel, D. W., and Torres, A. G. (2008). The effects of low-shear stress on Adherent-invasive *Escherichia coli*. *Environ. Microbiol.* 10, 1512–1525. doi: 10.1111/j.1462-2920.2008.01567.x
- Bettelheim, K. A. (1978). The sources of “OH” serotypes of *Escherichia coli*. *J. Hyg.* 80, 83–113. doi: 10.1017/S0022172400053420
- Beutin, L., Delannoy, S., and Fach, P. (2015a). Genetic diversity of the *fliCH19* genes encoding the flagellar antigen H19 of *Escherichia coli* and application to the specific identification of enterohemorrhagic *E. coli* (EHEC) O121:H19. *Appl. Environ. Microbiol.* 81, 4224–4230. doi: 10.1128/AEM.00591-15
- Beutin, L., Delannoy, S., and Fach, P. (2015b). Sequence variations in the flagellar antigen genes *fliCH25* and *fliCH28* of *Escherichia coli* and their use in identification and characterization of enterohemorrhagic *E. coli* (EHEC) O145:H25 and O145:H28. *PLoS ONE* 10:e0126749. doi: 10.1371/journal.pone.0126749
- Beutin, L., and Fach, P. (2014). “Detection of Shiga toxin-producing *Escherichia coli* from nonhuman sources and strain typing,” in *Microbiology Spectrum*, eds V. Sperandio and C. J. Hovde (Washington, DC: American Society for Microbiology), 1–23.
- Beutin, L., Krause, G., Zimmermann, S., Kaulfuss, S., and Gleier, K. (2004). Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. *J. Clin. Microbiol.* 42, 1099–1108. doi: 10.1128/JCM.42.3.1099-1108.2004
- Beutin, L., Kruger, U., Krause, G., Miko, A., Martin, A., and Strauch, E. (2008). Evaluation of major types of Shiga toxin 2e producing *Escherichia coli* present in food, pigs and in the environment as potential pathogens for humans. *Appl. Environ. Microbiol.* 74, 4806–4816. doi: 10.1128/AEM.00623-08
- Beutin, L., and Strauch, E. (2007). Identification of sequence diversity in the *Escherichia coli* *fliC* genes encoding flagellar types H8 and H40 and its use in typing of Shiga toxin-producing *E. coli* O8, O22, O111, O174, and O179 strains. *J. Clin. Microbiol.* 45, 333–339. doi: 10.1128/JCM.01627-06
- Bugarel, M., Beutin, L., Martin, A., Gill, A., and Fach, P. (2010). Microarray for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with Hemorrhagic Colitis and Hemolytic Uremic Syndrome in humans. *Int. J. Food Microbiol.* 142, 318–329. doi: 10.1016/j.ijfoodmicro.2010.07.010
- Clotilde, L. M., Salvador, A., Bernard, C., Lin, A., Lauzon, C., Muldoon, M., et al. (2015). Comparison of multiplex immunochemical and molecular serotyping methods for shiga toxin-producing *Escherichia coli*. *Foodborne Pathog. Dis.* 12, 118–121. doi: 10.1089/fpd.2014.1827
- Delannoy, S., Beutin, L., and Fach, P. (2013). Discrimination of enterohemorrhagic *Escherichia coli* (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. *J. Clin. Microbiol.* 51, 3257–3262. doi: 10.1128/JCM.01471-13
- Eaves-Pyles, T., Allen, C. A., Taormina, J., Swidsinski, A., Tutt, C. B., Jezek, G. E., et al. (2008). *Escherichia coli* isolated from a Crohn's disease patient adheres, invades, and induces inflammatory responses in polarized intestinal epithelial cells. *Int. J. Med. Microbiol.* 298, 397–409. doi: 10.1016/j.ijmm.2007.05.011

FUNDING

The project was partially financed by the French “joint ministerial program of R&D against CBRNE risks” (Grant number C1 7609- 2).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00135>

- Echeverria, P., Orskov, F., Orskov, I., and Plianbangchang, D. (1982). Serotypes of enterotoxigenic *Escherichia coli* in Thailand and the Philippines. *Infect. Immun.* 36, 851–856.
- Edwards, P. R., and Ewing, W. H. (1986). “The genus *Escherichia*,” in *Identification of Enterobacteriaceae*, 4th Edn., ed W. H. Ewing (New York, NY: Elsevier Science Publishing Co.), 93–136.
- Fairbrother, J. M., Nadeau, E., and Gyles, C. L. (2005). *Escherichia coli* in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. *Anim. Health Res. Rev.* 6, 17–39. doi: 10.1079/AHR2005105
- Fratamico, P. M., Bagi, L. K., Cray, W. C. Jr., Narang, N., Yan, X., Medina, M., et al. (2011). Detection by multiplex real-time polymerase chain reaction assays and isolation of Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. *Foodborne Pathog. Dis.* 8, 601–607. doi: 10.1089/fpd.2010.0773
- Frydendahl, K. (2002). Prevalence of serogroups and virulence genes in *Escherichia coli* associated with postweaning diarrhoea and edema disease in pigs and a comparison of diagnostic approaches. *Vet. Microbiol.* 85, 169–182. doi: 10.1016/S0378-1135(01)00504-1
- Geser, N., Stephan, R., Korczak, B. M., Beutin, L., and Hachler, H. (2012). Molecular identification of extended-spectrum-beta-lactamase genes from enterobacteriaceae isolated from healthy human carriers in Switzerland. *Antimicrob. Agents Chemother.* 56, 1609–1612. doi: 10.1128/AAC.05539-11
- Gomi, R., Matsuda, T., Fujimori, Y., Harada, H., Matsui, Y., and Yoneda, M. (2015). Characterization of pathogenic *Escherichia coli* in river water by simultaneous detection and sequencing of 14 virulence genes. *Environ. Sci. Technol.* 49, 6800–6807. doi: 10.1021/acs.est.5b00953
- Goulter, R. M., Gentle, I. R., and Dykes, G. A. (2010). Characterisation of curli production, cell surface hydrophobicity, autoaggregation and attachment behaviour of *Escherichia coli* O157. *Curr. Microbiol.* 61, 157–162. doi: 10.1007/s00284-010-9589-2
- Hampson, D. J., Fu, Z. F., Bettelheim, K. A., and Wilson, M. W. (1988). Management influences on the selective proliferation of two strains of haemolytic *Escherichia coli* in weaned pigs. *Epidemiol. Infect.* 100, 213–220. doi: 10.1017/S0950268800067340
- Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., et al. (2015a). A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res.* 22, 101–107. doi: 10.1093/dnares/dsu043
- Iguchi, A., Iyoda, S., Seto, K., Morita-Ishihara, T., Scheutz, F., and Ohnishi, M. (2015b). *Escherichia coli* O-Genotyping PCR: a Comprehensive and practical platform for molecular O Serogrouping. *J. Clin. Microbiol.* 53, 2427–2432. doi: 10.1128/JCM.00321-15
- Jakobsen, L., Spangholm, D. J., Pedersen, K., Jensen, L. B., Emborg, H. D., Agerso, Y., et al. (2010). Broiler chickens, broiler chicken meat, pigs and pork as sources of ExPEC related virulence genes and resistance in *Escherichia coli* isolates from community-dwelling humans and UTI patients. *Int. J. Food Microbiol.* 142, 264–272. doi: 10.1016/j.ijfoodmicro.2010.06.025
- Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by

- use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Johnson, J. R., Orskov, I., Orskov, F., Goulet, P., Picard, B., Moseley, S. L., et al. (1994). O, K, and H antigens predict virulence factors, carboxylesterase B pattern, antimicrobial resistance, and host compromise among *Escherichia coli* strains causing urosepsis. *J. Infect. Dis.* 169, 119–126. doi: 10.1093/infdis/169.1.119
- Johnson, J. R., Owens, K., Gajewski, A., and Kuskowski, M. A. (2005). Bacterial characteristics in relation to clinical course of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. *J. Clin. Microbiol.* 43, 6064–6072. doi: 10.1128/JCM.43.12.6064-6072.2005
- Johnson, J. R., Owens, K. L., Clabots, C. R., Weissman, S. J., and Cannon, S. B. (2006). Phylogenetic relationships among clonal groups of extraintestinal pathogenic *Escherichia coli* as assessed by multi-locus sequence analysis. *Microbes Infect.* 8, 1702–1713. doi: 10.1016/j.micinf.2006.02.007
- Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818
- Kaufmann, M., Zweifel, C., Blanco, M., Blanco, J. E., Blanco, J., Beutin, L., et al. (2006). *Escherichia coli* O157 and non-O157 Shiga toxin-producing *Escherichia coli* in fecal samples of finished pigs at slaughter in Switzerland. *J. Food Prot.* 69, 260–266.
- Martin, A., and Beutin, L. (2011). Characteristics of Shiga toxin-producing *Escherichia coli* from meat and milk products of different origins and association with food producing animals as main contamination sources. *Int. J. Food Microbiol.* 146, 99–104. doi: 10.1016/j.ijfoodmicro.2011.01.041
- Martinez-Medina, M., and Garcia-Gil, L. J. (2014). *Escherichia coli* in chronic inflammatory bowel diseases: an update on adherent invasive *Escherichia coli* pathogenicity. *World J. Gastrointest. Pathophysiol.* 5, 213–227. doi: 10.4291/wjgp.v5.i3.213
- Martinez-Medina, M., Naves, P., Blanco, J., Aldegue, X., Blanco, J. E., Blanco, M., et al. (2009). Biofilm formation as a novel phenotypic feature of adherent-invasive *Escherichia coli* (AIEC). *BMC Microbiol.* 9:202. doi: 10.1186/1471-2180-9-202
- Miko, A., Delannoy, S., Fach, P., Strockbine, N. A., Lindstedt, B. A., Mariani-Kurkdjian, P., et al. (2013). Genotypes and virulence characteristics of Shiga toxin-producing *Escherichia coli* O104 strains from different origins and sources. *Int. J. Med. Microbiol.* 303, 410–21. doi: 10.1016/j.ijmm.2013.05.006
- Minogue, T. D., Daligault, H. A., Davenport, K. W., Bishop-Lilly, K. A., Broomall, S. M., Bruce, D. C., et al. (2014). Complete genome assembly of *Escherichia coli* ATCC 25922, a serotype O6 reference strain. *Genome Announc.* 2:e00969-14. doi: 10.1128/genomeA.00969-14
- Nash, J. H., Villegas, A., Kropinski, A. M., Aguilar-Valenzuela, R., Konczy, P., Mascarenhas, M., et al. (2010). Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genomics.* 11:667. doi: 10.1186/1471-2164-11-667
- Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* 30, 371–403. doi: 10.1146/annurev.genet.30.1.371
- Olesen, B., Scheutz, F., Menard, M., Skov, M. N., Kolmos, H. J., Kuskowski, M. A., et al. (2009). Three-decade epidemiological analysis of *Escherichia coli* O15:K52:H1. *J. Clin. Microbiol.* 47, 1857–1862. doi: 10.1128/JCM.00230-09
- Orskov, F., and Orskov, I. (1984). “Serotyping of *Escherichia coli*,” in *Methods in Microbiology*, ed T. Bergan (London, UK: Academic Press), 43–112.
- Orskov, I., and Orskov, F. (1977). Special O:K:H serotypes among enterotoxigenic *E. coli* strains from diarrhea in adults and children. Occurrence of the CF (colonization factor) antigen and of hemagglutinating abilities. *Med. Microbiol. Immunol.* 163, 99–110. doi: 10.1007/BF02121825
- Orskov, I., and Orskov, F. (1985). *Escherichia coli* in extra-intestinal infections. *J. Hyg.* 95, 551–575. doi: 10.1017/S0022172400060678
- Pierard, D., Muyldermans, G., Moriau, L., Stevens, D., and Lauwers, S. (1998). Identification of new verocytotoxin type 2 variant B-subunit genes in human and animal *Escherichia coli* isolates. *J. Clin. Microbiol.* 36, 3317–3322.
- Ratiner, Y., Latto, I., Ulisko, I., Alekseeva, M., and Siitonen, A. (1995). Serological variety of flagellar antigen H1 in natural *Escherichia coli* population. *FEMS Immunol. Med. Microbiol.* 10, 125–131. doi: 10.1111/j.1574-695X.1995.tb00021.x
- Reid, S. D., Selander, R. K., and Whittam, T. S. (1999). Sequence Diversity of Flagellin (*flhC*) Alleles in Pathogenic *Escherichia coli*. *J. Bacteriol.* 181, 153–160.
- Reister, M., Hoffmeier, K., Krezdorn, N., Rotter, B., Liang, C., Rund, S., et al. (2014). Complete genome sequence of the gram-negative probiotic *Escherichia coli* strain Nissle 1917. *J. Biotechnol.* 187, 106–107. doi: 10.1016/j.jbiotec.2014.07.442
- Riley, L. W. (2014). Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* 20, 380–390. doi: 10.1111/1469-0691.12646
- Scheutz, F., Cheasty, T., Woodward, D., and Smith, H. R. (2004). Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new *E. coli* O groups that include Verocytotoxin-producing *E. coli* (VTEC): O176, O177, O178, O179, O180 and O181. *APMIS* 112, 569–584. doi: 10.1111/j.1600-0463.2004.apm1120903.x
- Scheutz, F., and Strockbine, N. A. (2005). “Genus I. *Escherichia*,” in *Bergey’s Manual of Systematic Bacteriology, 2nd Edn.*, eds G. M. Garrity, D. J. Brenner, N. R. Krieg, and J. T. Staley (New York, NY: Springer), 607–624.
- Scheutz, F., Teel, L. D., Beutin, L., Pierard, D., Buvens, G., Karch, H., et al. (2012). Multicenter evaluation of a sequence-based protocol for subtyping shiga toxins and standardizing stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12
- Shaheen, H. I., Khalil, S. B., Rao, M. R., Abu, E. R., Wierzbza, T. F., Peruski, L. F. Jr., et al. (2004). Phenotypic profiles of enterotoxigenic *Escherichia coli* associated with early childhood diarrhea in rural Egypt. *J. Clin. Microbiol.* 42, 5588–5595. doi: 10.1128/JCM.42.12.5588-5595.2004
- Singer, R. S. (2015). Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. *Front. Microbiol.* 6:28. doi: 10.3389/fmicb.2015.00028
- Stenutz, R., Weintraub, A., and Widmalm, G. (2006). The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol. Rev.* 30, 382–403. doi: 10.1111/j.1574-6976.2006.00016.x
- Tajima, F., and Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* 1, 269–285.
- Tschape, H., Bender, L., Ott, M., Wittig, W., and Hacker, J. (1992). Restriction fragment length polymorphism and virulence pattern of the veterinary pathogen *Escherichia coli* O139:K82:H1. *Zentralbl. Bakteriol.* 276, 264–272. doi: 10.1016/S0934-8840(11)80013-7
- Wang, L., Rothmund, D., Curd, H., and Reeves, P. R. (2000). Sequence diversity of the *Escherichia coli* H7 *flhC* genes: implication for a DNA-based typing scheme for *E. coli* O157:H7. *J. Clin. Microbiol.* 38, 1786–1790.
- Wang, L., Rothmund, D., Curd, H., and Reeves, P. R. (2003). Species-wide variation in the *Escherichia coli* flagellin (H-Antigen) gene. *J. Bacteriol.* 185, 2936–2943. doi: 10.1128/JB.185.9.2936-2943.2003
- Wang, Q., Wang, S., Beutin, L., Cao, B., Feng, L., and Wang, L. (2010). Development of a DNA microarray for detection and serotyping of enterotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 48, 2066–2074. doi: 10.1128/JCM.02014-09
- Welch, R. A., Burland, V., Plunkett, G. III, Redford, P., Roesch, P., Rasko, D., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17020–17024. doi: 10.1073/pnas.252529799
- Woudstra, C., Lambert, D., Anniballi, F., De Medici, D., Austin, J., and Fach, P. (2013). Genetic diversity of the flagellin genes of *Clostridium botulinum* groups I and II. *Appl. Environ. Microbiol.* 79, 3926–3932. doi: 10.1128/AEM.00686-13
- Zdzarski, J., Brzuszkiewicz, E., Wullt, B., Liesegang, H., Biran, D., Voigt, B., et al. (2010). Host imprints on bacterial genomes—rapid, divergent evolution in individual patients. *PLoS Pathog.* 6:e1001078. doi: 10.1371/journal.ppat.1001078

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

Copyright © 2016 Beutin, Delannoy and Fach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Rapid, Multiplexed Characterization of Shiga Toxin-Producing *Escherichia coli* (STEC) Isolates Using Suspension Array Technology

John M. Carter^{1*}, Andrew Lin², Laurie Clotilde³ and Matthew Lesho⁴

¹ Pacific West Area – Western Regional Research Center – Produce Safety and Microbiology Research, Agricultural Research Service, United States Department of Agriculture, Albany, CA, USA, ² ORA/PA-FO/SAN-LB – Office of Global Regulatory Operations and Policy – Oceans, Reefs & Aquariums – Food and Drug Administration, United States Department of Health and Human Services, Alameda, CA, USA, ³ MagArray, Inc., Milpitas, CA, USA, ⁴ Luminex Corporation, Austin, TX, USA

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Arun K. Bhunia,
Purdue University, USA
Yanhong Liu,
United States Department of
Agriculture, USA

*Correspondence:

John M. Carter
mark@safetraces.com

Specialty section:

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

Received: 15 January 2016

Accepted: 18 March 2016

Published: 20 May 2016

Citation:

Carter JM, Lin A, Clotilde L
and Lesho M (2016) Rapid,
Multiplexed Characterization of Shiga
Toxin-Producing *Escherichia coli*
(STEC) Isolates Using Suspension
Array Technology.
Front. Microbiol. 7:439.
doi: 10.3389/fmicb.2016.00439

Molecular methods have emerged as the most reliable techniques to detect and characterize pathogenic *Escherichia coli*. These molecular techniques include conventional single analyte and multiplex PCR, PCR followed by microarray detection, pulsed-field gel electrophoresis (PFGE), and whole genome sequencing. The choice of methods used depends upon the specific needs of the particular study. One versatile method involves detecting serogroup-specific markers by hybridization or binding to encoded microbeads in a suspension array. This molecular serotyping method has been developed and adopted for investigating *E. coli* outbreaks. The major advantages of this technique are the ability to simultaneously serotype *E. coli* and detect the presence of virulence and pathogenicity markers. Here, we describe the development of a family of multiplex molecular serotyping methods for Shiga toxin-producing *E. coli*, compare their performance to traditional serotyping methods, and discuss the cost-benefit balance of these methods in the context of various food safety objectives.

Keywords: *E. coli*, Shiga toxin, immunoassay, PCR, microbead

INTRODUCTION/BACKGROUND

In the U.S., Shiga toxin-producing *Escherichia coli* (STEC) represent a significant public health concern causing approximately 175,000 illnesses annually (Scallan et al., 2011). The CDC estimates the annual U.S. costs for acute care of STEC patients to be \$1–2 billion (Paton and Paton, 2000 Science and Medicine p 28–37). Furthermore, STEC-related recalls are expensive for the food industry. For example, the outbreak of *E. coli* O157 associated with spinach from California's Salinas Valley in 2006 cost farmers up to \$200 million (Arnade et al., 2010). Although *E. coli* O157 remains the most common STEC serogroup in the U.S. (Centers for Disease Control and Prevention [CDC], 2011), over 200 additional *E. coli* serogroups exist (Hussein and Bollinger, 2005). Those non-O157 STEC are responsible for over 60% of the STEC infections (Scallan et al., 2011). The clinical manifestations of STEC infections range from mild watery diarrhea to severe complications of hemorrhagic colitis (HC), hemolytic uremic syndrome (HUS), and even death (Levine et al., 1987; Gyles, 2007). Since not all STEC appear to cause illness, distinguishing pathogenic STEC from the ones that do not pose a health risk remains a current challenge for regulatory agencies worldwide. Certain STEC serogroups are more highly correlated with severe

illness than others. For example, STEC O26, O45, O103, O111, O121, and O145 are referred to as the “Big Six” and account for 75–80% of non-O157 STEC isolations in clinical samples in the U.S. (Gould et al., 2013). In addition, STEC O91, O113, and O128 have also been previously reported to cause HC and HUS (Johnson et al., 2006), and an STEC O104-caused outbreak in Germany in 2011 sickened 3816 individuals, making it one of the largest HUS outbreaks ever reported (Muniesa et al., 2012). Thus rapidly identifying STEC belonging to those serogroups is as important for protecting consumer health as early diagnosis of STEC infection for determining the proper treatment (Gould et al., 2009). Additionally, such rapid identification can alert public health officials that an outbreak has occurred and aid in matching clinical, food, and environmental isolates when attempting to trace back to the source of contamination.

Traditional serotyping employs O-specific antisera, usually in a slide agglutination format. These methods are quite simple, robust, and rapid. However, they can be both labor-intensive and time-consuming for large numbers of isolates, and may lead to ambiguous results because of variability in antisera production and a lack of standard methodology. Whereas slide agglutination is also relatively inexpensive when only a few of the most common antigens are tested, maintaining a complete set of hundreds of antisera reagents for O serotyping is expensive and most often left to central reference laboratories adding to the time of analysis (Dijkshoorn et al., 2001). Currently, there is a need for faster tests. One solution is to allow a food sample to be probed simultaneously for specific STEC serogroups, which can then be extracted from that food sample, thus completing detection and isolation of STEC in foods within 24 h. Suspension array technology using microbeads has been useful in filling this need by enabling development of rapid, high-throughput adaptable assays for identifying clinically relevant STEC O-serogroups. These microbead-based assays can theoretically analyze up to 500 targets in a single reaction, so targets can be added or removed as needed. For example, a 7-plex immunoassay was developed by Clotilde et al. (2013) to identify O serogroups O26, O45, O103, O111, O121, O145, and O157. The microbeads were also capable of binding to live cells, which aides in isolating pathogenic organisms for further characterization studies (Clotilde et al., 2013). Another example is a 13-plex PCR-based suspension assay to identify 11 O-serogroups: O26, O45, O91, O103, O104, O111, O113, O121, O128, O145, and O157; and two virulence factors: *eae* and *aggR* (Feng et al., 2015). These two assays are fast (<4 h), high throughput (96 well format), and were able to identify STEC serogroups more accurately than traditional serotyping (Clotilde et al., 2015).

SUSPENSION ARRAY TECHNOLOGY

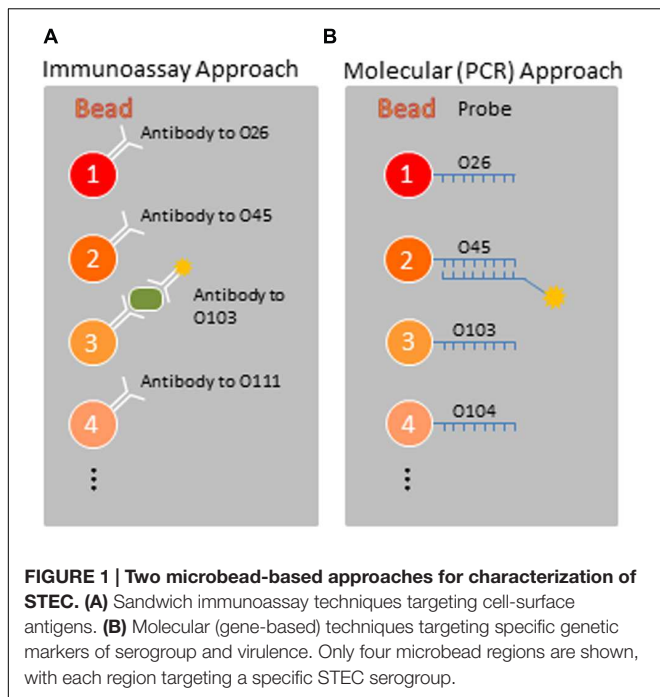
Microbead-based suspension array technology has emerged as a standard method for simultaneously detecting multiple biological analytes from one sample and has enabled a wide variety of applications in life science research, clinical diagnostics, food safety, and biodefense (McBride et al., 2003; Toro et al., 2013; Sun et al., 2014; Di Cristanziano et al., 2015; Khalifian et al.,

2015; Silbereisen et al., 2015; Zhang et al., 2015; Kong et al., 2016). Commercial assay kits are available for cytokine profiling, infectious disease diagnostics, genotyping for inherited diseases, food pathogen typing, organ transplant compatibility testing, and many more. Many of these kits are built on the Luminex® xMAP® multiplexing platform, which utilizes fluorescent dyes to create sets of microbeads with unique spectral identities. Unique capture molecules are coupled to each set of microbeads, which then capture different analytes of interest. After binding to a detector molecule, which subsequently binds to a fluorescent reporter, these microbeads are read in a flow cytometer such as the Luminex® 200™ or FLEXMAP 3D® instrument, or in a bead imager such as the Luminex MAGPIX® instrument to determine both the presence and quantity of analyte(s) in the sample. The Luminex systems are capable of detecting up to 50, 100, or 500 different analytes from one sample in the MAGPIX, Luminex 200, or FLEXMAP 3D instruments, respectively. Some advantages of suspension array platforms include fast kinetics due to mobile capture surfaces, broad dynamic range (>3 logs), and high precision due to multiple independent measurements on many microbeads for each analyte in the sample. Open platform suspension array systems also enable researchers and other biomedical professionals to design and build tests for their specific applications. In the area of food safety discussed here, an important aspect of suspension array technology is the ability to detect DNA, RNA, and protein targets with the same system.

IMMUNOASSAY/PROTEIN BASED SEROTYPING

Inoculation of *E. coli* produces a strong immune response, which targets immunodominant antigens, such as lipopolysaccharide (O-antigens) and flagellum (H-antigens). The best-known *E. coli* serotype is O157:H7. Strains sharing the O157 O-antigen (i.e., members of the O157 serogroup) also tend to share pathogenic phenotypes. For example, most strains belonging to the O157 serogroup exhibit prophages that code for Shiga toxin (Stx). Other serotypes may also carry Stx-encoding phage, and in the US seven STEC serogroups are considered adulterants in foods: O26, O111, O103, O121, O45, O145, and O157. Although there are particular strains within these serogroups that do not produce Stx, all are grounds for regulatory actions, including product recalls. Thus the food industries have considerable interest in *E. coli* serogroup testing.

Identification of specific *E. coli* serotypes is important for tracing infections to their environmental source. Outbreaks were formerly detected and defined solely by the serogroup of pathogens, and serotyping still provides a rapid and inexpensive means for preliminary characterization. Until recently the “gold standard” for serotyping was slide agglutination with O and H antigen specific antisera (Orskov et al., 1977; Machado et al., 2000). Now, alternative immunochemical assay formats are also available, including our Luminex-format microbead suspension array as a solid phase for multiplex typing of various STEC serogroups. (See **Figure 1A**).



Our assay architecture is similar to a sandwich ELISA, starting with an antibody (Ab) that is bound to a solid phase support. In a typical ELISA, this “capture Ab” is non-covalently adsorbed to a microplate well, but we use covalent attachment. When the target is added, the capture antibody can bind and pull its antigen out of the solution phase. Next the detector Ab is added, which binds to the antigen to create the Ab sandwich configuration. Finally the detector Ab is detected by means of a probe. In ELISA, that probe is labeled with an enzyme. In our assay, the probe is labeled with phycoerythrin, a bright fluorophore. Because bacteria exhibit many copies of O-antigen, the same Ab may be used for both capture and detection.

To generate the assay reagents, magnetic microbeads were covalently coupled to capture Ab according to the instructions provided with the BioRad Amine Coupling kit (BioRad, Hercules, CA, USA), using an amount of Ab based on a preliminary microplate ELISA data (1–10 µg/mL). This reaction is a common two-step carbodiimide protocol with N-hydroxysulfosuccinimide. Coupling was confirmed using phycoerythrin-labeled anti-species Ab. Detector Ab were biotinylated with the EZ-Link Sulfo-NHS-Long Chain-Biotin kit (Pierce, Rockford, IL, USA), according to the manufacturer’s instructions. Coupled microbeads and detector Ab were stable at 2–8°C for up to 1 year.

Samples were generated by blending foods at a 1:10 dilution into Brain Heart Infusion broth. A 1mL aliquot was then spiked with ~10 CFU of *E. coli* and incubated overnight shaking at 25 or 37°C. Appropriate BSL2 precautions were used when handling live pathogens. For example, samples were spiked after blending, and cultures were handled in screw-cap tubes, to reduce aerosolization.

For the assay, incubations were all 1 h at room temperature, in black microplates, swirling at 100 rpm. Washes were all threefold, using phosphate-buffered saline (PBS), pH 7.4, containing 0.05% Tween 20 (PBS-T) and a Bio-Plex Pro Wash Station (BioRad). For each sample, a 100 µL aliquot was combined with 5000 of each type of microbead, then incubated and washed. The microbeads were then resuspended in 100 µL of a cocktail containing 4 µg/mL of each biotinylated detector Ab, then incubated and washed. The microbeads were then resuspended in 100 µL of 4 µg/mL streptavidin labeled with R-phycoerythrin (SAPE), incubated and washed. The microbeads were finally resuspended in 100 µL PBS and then analyzed on Luminex 100 or MAGPIX instruments.

For these experiments, we used overnight cultures, and we were not concerned with assay sensitivity. We typically recovered 50–90% of microbeads from samples, depending on the matrix. We found that even 100 microbeads (instead of 5000) were sufficient for robust and reproducible signals. In validation experiments performed with three strains of each of the seven listed STEC serogroups, the assay gave 100% sensitivity and specificity (Clotilde et al., 2013). When we expanded the validation to include 161 environmental STEC strains our Luminex immunoassay missed only one strain of O157. In this latter experiment we compared performance of our assay to standard assays and a microbead-based PCR serotyping assay (Clotilde et al., 2015). This comparative study (Table 1) suggested that our Luminex immunoassay also missed 10 strains of O111 and O128. However, these targets were not included in our 7-plex immunoassay.

We have now further expanded our assay to include the 10 most common STEC in the US: O26, O45, O103, O111, O121, O145, and O157, plus O91, O113, and O128. We still observe excellent sensitivity and specificity. We believe that STEC will continue to evolve additional pathogenic serogroups in the future. Regulatory agencies currently base their actions on the serogroup of such emerging adulterants, but there is consensus toward surveillance of virulence factors, rather than O-antigens. We already have working immunoassays for Stx (Clotilde et al., 2011), and we plan to add intimin, a virulence factor involved in *E. coli* attachment.

MOLECULAR SEROTYPING

The O antigen, a polymer of repeating oligosaccharides, is a component of the lipopolysaccharide of gram negative bacteria, and is used for serotyping *E. coli* into O serogroups. The *wzx* flippase, and *wzy* polymerase genes code for proteins involved in making the O antigen oligosaccharide, have proven to be O serogroup specific, and thus excellent targets for O serogroup specific PCR assays (Lin et al., 2011b). These genes were used to develop a multiplex suspension array that can identify 11 STEC O serogroups: O26, O45, O91, O103, O104, O111, O113, O121, O128, O145, and O157 (Lin et al., 2013) (Figure 1B).

Primer and probe sequences from each of the 11 O serogroup *wzx* or *wzy* genes were identified. Primers were designed with one biotinylated primer, and one unlabeled primer to result

TABLE 1 | List of disagreements between serotyping methods.

Strain #	Conventional	Luminex		Additional serotyping	
		PCR	Antibody	Statens	Remel
4	OUT	O157	O157		O157
5	OUT	O157	O157		O157
8	OUT	O157	O157		O157
11	OUT	O157	O157		O157
60	OUT	O157	O157		O157
83	OUT	O157	O157		O157
116	OUT	O157	O157		O157
126	OUT	O157	O157		O157
120	OUT	O45	O45		
25	OUT	O111	–	O111	
32	OUT	O111	–	O111	
69	O127	O128	–	O128	
71	O125	O128	–	O128	
72	O111	O128	–	O128	
73	O125	O128	–	O128	
74	O127	O128	–	O128	
75	O127	O128	–	O128	
77	O127	O128	–	O128	
70	O166	O128 and O121	O121	O128 and O121	
86	O125	O157	O157		O157
105	O158	O157	O157		O157
88	O137	O111	O111	O111	
85	O157	O157	–		

OUT, other untypeable.

in a biotinylated amplicon. A single primer mix with primers for all targets was prepared. PCR reactions were carried out according to Lin et al. (2013). Signal to noise was improved with asymmetric PCR amplification with concentrations of biotinylated: unlabeled primer at a ratio of 5:1. Probe sequences complimentary to the biotinylated strand were conjugated to Luminex MagPlex® microbeads according to the manufacturer's instructions (Luminex Corporation, 2007). After multiplex PCR, amplicons were hybridized to labeled microbeads and incubated with SAPE. Reactions were then analyzed with a Luminex compatible cytometry-type instrument to interrogate each uniquely colored microbead and detect the amount of reporter molecule. Median fluorescent intensities (MFI's) were calculated for each analyte, and a signal to background ratio was determined, where background is the MFI measured using one or more wells containing all ingredients except for template DNA. Signal to background ratio was calculated using Bio-Plex Manager software as (Sample MFI-Background MFI)/Background MFI. Samples are considered positive when signal to background ratio is greater than 5.0.

The PCR based Luminex suspension array for O serotyping has proven to be accurate, robust, and adaptable. A panel of 114 STEC isolates were all correctly identified, while 46 non-STEC and non- *E. coli* yielded no false positives (Lin et al., 2011a). In a multi-laboratory study of blind samples involving nine laboratories, a total of 492 isolates were identified correctly

out of 495 analyzed (99.4% accuracy) (Lin et al., 2013). Another advantage of the suspension array system and of PCR based suspension arrays is the flexibility to add or remove targets from the array as needed. For instance, the 11-plex suspension array to identify 11 STEC O serogroups was recently modified to include STEC attachment factors that may be important markers of pathogenic potential. Primers and probes for the *eae* intimin gene, and the *aggR* regulator of the enteroaggregative phenotype have been added to the 11-plex O serotyping array to comprise a 13-plex STEC suspension array (Feng et al., 2015).

Suspension array technology allows for rapid, accurate, high throughput analysis of STEC isolates. In addition to identifying the most clinically relevant STEC O serogroups, the 13-plex suspension array is able to detect the presence of adherence factors that are associated with HUS, allowing regulatory and public health labs to determine pathogenic potential. The addition of the virulence factors to the suspension array also illustrates the flexibility of suspension array technology. While the big six STEC O serogroups and the O91, O128, O113, O104 additional O serogroups are presently the most concerning, new emerging O serogroups could be added to the suspension array. Additional targets could be added to the array, if other genes prove to be more highly correlated with illness. For instance, the H flagellar antigen is a useful phylogenetic marker (Ju et al., 2012) and could be added to the array to indicate the full O and H serotype. Other putative virulence factors that could be included

in an STEC virulence profile include the plasmid-encoded enterohemolysin (*ehxA*), STEC autoagglutinating adhesin (*Saa*), extracellular serine protease (*espP*), long polar fimbria (*lpf*), and a bifunctional catalase-peroxidase (*katP*) (Law, 2000; Etcheverria and Padola, 2013). Another area of future study is to evaluate the effectiveness of the suspension array in screening food, environmental, and clinical samples. A preliminary study of artificially inoculated fresh produce and raw milk resulted in over 90% agreement between the suspension array and qPCR screening for *stx1* and/or *stx2* genes (Kase et al., 2015). Further improvements such as including an internal amplification control would be useful especially when screening enrichments to ensure that PCR inhibition does not cause false negative results.

DISCUSSION

Both of these suspension array assays represent improvements in speed and accuracy for detection and characterization of STEC pathogens in foods. A unique feature of the microbead-based immunoassay is that the magnetic microbeads pull down live pathogens, concentrating them. We have found that positive samples are easily cultured from the material remaining in the assay microplate (Clotilde et al., 2011). This facilitates further characterization, e.g., via molecular experiments. Other possibilities include use of an orthogonal detection scheme, other than SAPE. One group has combined immunomagnetic pull-down on Luminex microbeads with a modified fluorescence in situ hybridization (FISH) detection method (Stroot et al., 2012). This capture antibody-targeted FISH assay (CAT-FISH) method provides sensitivity equal to the Luminex immunoassay with the possibility of additional specificity from the nucleic acid-based FISH reporter. Among the many advantages of these suspension arrays is the ability to quickly upgrade and enhance assays as new genomic information becomes available for other pathotypes. Beyond the addition of more informative targets, other potential enhancements for these assays include workflow simplification, improved reagent stability through lyophilization or stable, engineered antibodies, and the ability to interrogate single cells with the assay, which would enable the characterization of mixed populations of STEC.

These assays fit on the spectrum of STEC characterization techniques between single-plex PCR assays for rapid *E. coli* screening of food samples and Next-Generation Sequencing (NGS) techniques. NGS is benefitting assay developers by providing information about novel molecular targets to more

completely characterize STEC samples. While NGS is a powerful tool in microbiological research and outbreak investigations, there is still a need for low-complexity multiplex assays to make specific determinations about prioritized serogroups. One of the most significant drawbacks to fluorescent microbead-based methods is the cost of reagents, specifically the microbeads themselves. If we test 100 samples with a 10-plex assay, then 1000 assays are performed, and the cost per assay is acceptable. However, if half of those samples are identified as O157, then a more cost-effective work flow might begin with slide agglutination for all samples, using a single-plex O157 assay, followed by work-up of the non-O157 samples using a more expensive multiplex assay. The tradeoff of this workflow is that the time taken to do a full determination of the non-O157 samples is increased significantly. The suspension array technology still remains a versatile platform that benefits research, industrial, and regulatory labs by enabling a variety of protein and genetic analyses for many foodborne pathogens of interest (e.g., sets of assays to identify and characterize *Salmonella*, STEC, and foodborne toxins).

AUTHOR CONTRIBUTIONS

JC Drafted and edited immunoassay/protein based serotyping section and discussion. He and his lab staff conceived, developed, built, and tested the assays that resulted in the data in **Table 1**. AL Drafted and edited the Introduction and Molecular Serotyping sections. He conceived and developed the molecular serotyping assays described herein, including the enhanced assay with markers of pathogenic potential. LC Provided significant editorial contribution to all sections. Collaborated with JC and AL in building and testing the assays described. ML Outlined the article, drafted abstract and suspension array sections and formulated key messages in the discussion section. Luminex Corporation currently manufactures *E. coli* Serogroup Identification assays.

ACKNOWLEDGMENTS

U.S. Department of Agriculture work was funded under National Program 108, Food Safety, CRIS 5325-42000-043-00D. The authors would also like to thank the FDA San Francisco Laboratory, FDA Office of Regulatory Science, and CFSAN Office of Regulatory Science for their support of this research.

REFERENCES

- Arnade, C., Calvin, L., and Kuchler, F. (2010). Consumers' response to the 2006 foodborne illness outbreak linked to spinach. *Amber Waves*. Available at: <http://www.ers.usda.gov/amber-waves/2010-march/consumers%E2%80%99-response-to-the-2006-foodborne-illness-outbreak-linked-to-spinach.aspx#.VzHyxvkrKU1>
- Centers for Disease Control and Prevention [CDC] (2011). *OutbreakNet, Foodborne Outbreak Online Database*. Available at: <http://wwwn.cdc.gov/foodborneoutbreaks/> [Accessed on September 10, 2012].
- Clotilde, L. M., Bernard, C. I. V., Hartman, G., Lau, D. K., and Carter, J. M. (2011). Microbead-Based immunoassay for simultaneous detection of Shiga toxins and isolation of *Escherichia coli* O157 in foods. *J. Food Prot.* 74, 373–379. doi: 10.4315/0362-028X.JFP-10-344
- Clotilde, L. M., Bernard, C. I. V., Salvador, A., Lin, A., Lauzon, C. R., Muldoon, M., et al. (2013). A 7-plex microbead-based immunoassay for serotyping Shiga toxin-producing *Escherichia coli*. *J. Microbiol. Methods* 92, 226–230. doi: 10.1016/j.mimet.2012.11.023
- Clotilde, L. M., Salvador, A., Bernard, C. I. V., Lin, A., Lauzon, C., Muldoon, M., et al. (2015). Comparison of multiplex immunochemical and molecular

- serotyping methods for Shiga toxin-producing *Escherichia coli*. *Foodborne Path. Dis.* 12, 118–121. doi: 10.1089/fpd.2014.1827
- Di Cristanziano, V., Timmen-Wego, M., Lübke, N., Kaiser, R., Pfister, H., Di Cave, D., et al. (2015). Application of luminex gastrointestinal pathogen panel to human stool samples from côte d'ivoire. *J. Infect. Dev. Ctries* 9, 884–889. doi: 10.3855/jidc.6460
- Dijkshoorn, L., Towner, K. J., and Struelens, M. (2001). *New Approaches for the Generation and Analysis of Microbial Typing Data*. Amsterdam: Elsevier Science B. V.
- Etcheverria, A. I., and Padola, N. L. (2013). Shiga toxin-producing *Escherichia coli* Factors involved in virulence and cattle colonization. *Virulence* 4, 366–372. doi: 10.4161/viru.24642
- Feng, P., Weagant, S. D., and Jinneman, K. (2015). *Bacteriological Analytical Manual Chapter 4a Diarrheagenic Escherichia coli*. Available at: <http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm070080.htm>
- Gould, L. H., Bopp, C., Strockbine, N., Atkinson, R., Baselski, V., Body, B., et al. (2009). Recommendations for diagnosis of shiga toxin-producing *Escherichia coli* infections by clinical laboratories. *MMWR Recomm. Rep.* 58, 1–14.
- Gould, L. H., Mody, R. K., Ong, K. L., Clogher, P., Cronquist, A. B., Garman, K. N., et al. (2013). Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000–2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Path. Dis.* 10, 453–460. doi: 10.1089/fpd.2012.1401
- Gyles, C. L. (2007). Shiga toxin-producing *Escherichia coli*: an overview. *J. Anim. Sci.* 85, E45–E62. doi: 10.2527/jas.2006-508
- Hussein, H. S., and Bollinger, L. M. (2005). Prevalence of Shiga toxin-producing *Escherichia coli* in beef cattle. *J. Food Prot.* 68, 2224–2241.
- Johnson, K. E., Thorpe, C. M., and Sears, C. L. (2006). The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. *Clin. Infect. Dis.* 43, 1587–1595. doi: 10.1086/509573
- Ju, W., Cao, G., Rump, L., Strain, E., Luo, Y., Timme, R., et al. (2012). Phylogenetic analysis of Non-O157 shiga toxin-producing *Escherichia coli* strains by whole-genome sequencing. *J. Clin. Microbiol.* 50, 4123–4127. doi: 10.1128/JCM.02262-12
- Kase, J. A., Maounounen-Laasri, A., and Lin, A. (2015). “Rapid identification of eleven STEC O serogroups from fresh produce and raw milk enrichments by luminex bead based suspension array,” in *Proceeding of the 2015 VTEC Meeting*, Boston, MA.
- Khalifian, S., Raimondi, G., and Brandacher, G. (2015). The use of luminex assays to measure cytokines. *J. Invest. Derm.* 135:e31. doi: 10.1038/jid.2015.36
- Kong, W., Lib, Y., Cheng, S., Yan, C., An, S., Dong, Z., et al. (2016). Luminex xMAP combined with Western blot improves HIV diagnostic sensitivity. *J. Virol. Methods* 227, 1–5. doi: 10.1016/j.jviromet.2015.10.007
- Law, D. (2000). Virulence Factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*. *J. Appl. Microbiol.* 88, 729–745. doi: 10.1046/j.1365-2672.2000.01031.x
- Levine, M. M., Xu, J. G., Kaper, J. B., Lior, H., Prado, V., Tall, B., et al. (1987). A DNA probe to identify enterohemorrhagic *Escherichia coli* of O157:H7 and other serotypes that cause hemorrhagic colitis and hemolytic uremic syndrome. *J. Infect. Dis.* 156, 175–182. doi: 10.1093/infdis/156.1.175
- Lin, A., Kase, J. A., Moore, M. M., Son, I., Tran, N., Clotilde, L. M., et al. (2013). Multilaboratory validation of a Luminex microbead-based suspension array for the identification of the 11 most clinically relevant Shiga toxin-producing *Escherichia coli* O serogroups. *J. Food Prot.* 76, 867–870. doi: 10.4315/0362-028X.JFP-12-468
- Lin, A., Nguyen, L., Lee, T., Clotilde, L. M., Kase, J. A., Son, I., et al. (2011a). Rapid O serogroup identification of the ten most clinically relevant STECs by Luminex microbead-based suspension array. *J. Microbiol. Methods* 87, 105–110. doi: 10.1016/j.mimet.2011.07.019
- Lin, A., Sultan, O., Lau, H. K., Wong, E., Hartman, G., and Lauzon, C. R. (2011b). O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. *Food Microbiol.* 28, 478–483. doi: 10.1016/j.fm.2010.10.007
- Luminex Corporation (2007). *Sample Protocol for Carbodiimide Coupling of Amine-Modified Oligonucleotides to MagPlex™-C Magnetic Carboxylated Microspheres*. Available at: <http://www.luminexcorp.com/prod/groups/public/documents/lmnxcorp/magnetic-oligo-coupling-protoc.pdf>
- Machado, J., Grimont, F., and Grimont, P. A. (2000). Identification of *Escherichia coli* flagellar types by restriction of the amplified *flhC* gene. *Res. Microbiol.* 151, 535–546. doi: 10.1016/S0923-2508(00)00223-0
- McBride, M. T., Gammon, S., Pitesky, M., O'Brien, T. W., Smith, T., Aldrich, J., et al. (2003). Multiplexed liquid arrays for simultaneous detection of simulants of biological warfare agents. *Anal. Chem.* 75, 1924–1930. doi: 10.1021/ac026379k
- Muniesa, M., Hammerl, J. A., Hertwig, S., Appel, B., and Brüssow, H. (2012). Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Appl. Environ. Microbiol.* 78, 4065–4073. doi: 10.1128/AEM.00217-12
- Orskov, I., Orskov, F., Jann, B., and Jann, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol. Rev.* 41, 667–710.
- Paton, J. C., and Paton, A. W. (2000). Shiga toxigenic *Escherichia coli* infections. *Sci. Med.* 28–37.
- Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M., Roy, S. L., et al. (2011). Foodborne illness acquired in the United States—major pathogens. *Emerging Infect. Dis.* 17, 7–15. doi: 10.3201/eid1701.091101p1
- Silbereisen, A., Tamborini, M., Wittwer, M., Nadia Schürch, N., and Pluschke, G. (2015). Development of a bead-based Luminex assay using lipopolysaccharide specific monoclonal antibodies to detect biological threats from *Brucella species*. *BMC Microbiol.* 15:198. doi: 10.1186/s12866-015-0534-1
- Stroot, J. A., Leach, K. M., Stroot, P. G., and Lim, D. V. (2012). Capture antibody targeted fluorescence in situ hybridization (CAT-FISH): Dual labeling allows for increased specificity in complex samples. *J. Microbiol. Methods* 88, 275–284. doi: 10.1016/j.mimet.2011.12.009
- Sun, Z., Yuan Peng, Y., Zhang, M., Wang, K., Bai, J., Li, X., et al. (2014). Simultaneous and highly sensitive detection of six different foodborne pathogens by high-throughput suspension array technology. *Food Control* 40, 300–309. doi: 10.1016/j.foodcont.2013.11.028
- Toro, M., Najjar, M. B., Ju, W., Brown, E., Zhao, S., and Meng, J. (2013). Molecular Serogrouping of Shiga Toxin-Producing *Escherichia coli* Using Suspension Array. *Foodborne Path. Dis.* 10, 478–480. doi: 10.1089/fpd.2012.1352
- Zhang, H., Morrison, S., and Tang, Y. W. (2015). Multiplex polymerase chain reaction tests for detection of pathogens associated with gastroenteritis. *Clin. Lab. Med.* 35, 461–486. doi: 10.1016/j.cll.2015.02.006

Disclaimer: This presentation is not intended to be or appear as an endorsement, either directly, or indirectly, by the U.S. Government of any product or service described in this presentation. The U.S. Government does not directly or indirectly endorse any product or service presented or described in this presentation. The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, SW, Washington, DC 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA funding was administered through the Agricultural Research Service, National Program 108 Food Safety.

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

Copyright © 2016 Carter, Lin, Clotilde and Lesho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Six Novel O Genotypes from Shiga Toxin-Producing *Escherichia coli*

Atsushi Iguchi^{1*}, Sunao Iyoda², Kazuko Seto³, Hironobu Nishii¹, Makoto Ohnishi², Hirohisa Mekata^{4,5}, Yoshitoshi Ogura⁶ and Tetsuya Hayashi⁶

¹ Department of Animal and Grassland Sciences, Faculty of Agriculture, University of Miyazaki, Miyazaki, Japan,

² Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan, ³ Division of Bacteriology, Osaka Prefectural Institute of Public Health, Osaka, Japan, ⁴ Organization for Promotion of Tenure Track, University of Miyazaki, Miyazaki, Japan, ⁵ Center for Animal Disease Control, University of Miyazaki, Miyazaki, Japan, ⁶ Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan

OPEN ACCESS

Edited by:

Pina Fratamico,
United States Department
of Agriculture – Agricultural Research
Service, USA

Reviewed by:

Beatriz Quiñones,
United States Department
of Agriculture – Agricultural Research
Service, USA
Edward G. Dudley,
Pennsylvania State University, USA

*Correspondence:

Atsushi Iguchi
iguchi@med.miyazaki-u.ac.jp

Specialty section:

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

Received: 18 February 2016

Accepted: 06 May 2016

Published: 20 May 2016

Citation:

Iguchi A, Iyoda S, Seto K, Nishii H,
Ohnishi M, Mekata H, Ogura Y
and Hayashi T (2016) Six Novel O
Genotypes from Shiga
Toxin-Producing *Escherichia coli*.
Front. Microbiol. 7:765.
doi: 10.3389/fmicb.2016.00765

Serotyping is one of the typing techniques used to classify strains within the same species. O-serogroup diversification shows a strong association with the genetic diversity of O-antigen biosynthesis genes. In a previous study, based on the O-antigen biosynthesis gene cluster (O-AGC) sequences of 184 known *Escherichia coli* O serogroups (from O1 to O187), we developed a comprehensive and practical molecular O serogrouping (O genotyping) platform using a polymerase chain reaction (PCR) method, named *E. coli* O-genotyping PCR. Although, the validation assay using the PCR system showed that most of the tested strains were successfully classified into one of the O genotypes, it was impossible to classify 6.1% (35/575) of the strains, suggesting the presence of novel O genotypes. In this study, we conducted sequence analysis of O-AGCs from O-genotype untypeable Shiga toxin-producing *E. coli* (STEC) strains and identified six novel O genotypes; OgN1, OgN8, OgN9, OgN10, OgN12 and OgN31, with unique wzx and/or wzy O-antigen processing gene sequences. Additionally, to identify these novel O-genotypes, we designed specific PCR primers. A screen of O genotypes using O-genotype untypeable strains showed 13 STEC strains were classified into five novel O genotypes. The O genotyping at the molecular level of the O-AGC would aid in the characterization of *E. coli* isolates and will assist future studies in STEC epidemiology and phylogeny.

Keywords: *E. coli*, O serogroup, genotyping techniques, PCR, STEC

INTRODUCTION

Serotyping is a standard method for subtyping of *Escherichia coli* strains in taxonomical and epidemiological studies (Orskov and Orskov, 1984). In particular, the identification of strains of the same O serogroup is essential in outbreak investigations and surveillance for identifying the diffusion of a pathogenic clone (Frank et al., 2011; Luna-Gierke et al., 2014; Terajima et al., 2014; Heiman et al., 2015). Thus far, the World Health Organization Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella*, which is based at the Statens Serum Institut (SSI) in Denmark¹, has recognized 185 *E. coli* O serogroups. These are designated O1 to O188 (publication of O182 to O188 is pending) and include three pairs of subgroups, O18ab/ac, O28ab/ac, and O112ab/ac; and six missing numbers, O31, O47, O67, O72, O93, and O122 (Orskov and Orskov, 1992; Scheut et al., 2004).

¹ <http://www.ssi.dk/English.aspx>

O-serogroup diversification shows a strong association with the genetic diversity of O-antigen biosynthesis genes. In *E. coli*, the genes required for O-antigen biosynthesis are clustered at a chromosomal locus flanked by the colanic acid biosynthesis gene cluster (*wca* genes) and the histidine biosynthesis (*his*) operon. Sequence comparisons of O-antigen biosynthesis gene clusters (O-AGCs) indicate a variety of genetic structures (DebRoy et al., 2011a). In particular, sequences from O-antigen processing genes (*wzx/wzy* and *wzm/wzt*) located on the O-AGCs are highly variable and can be used as gene markers for the identification of O serogroups via molecular approaches. So far, several studies have reported genetic methodologies allowing rapid and low-cost O-typing of isolates (Coimbra et al., 2000; Beutin et al., 2009; Bugarel et al., 2010; Wang et al., 2010, 2014; DebRoy et al., 2011b; Fratamico and Bagi, 2012; Quiñones et al., 2012; Geue et al., 2014). In a previous study (Iguchi et al., 2015a), we analyzed the O-AGC sequences of 184 known *E. coli* O serogroups (from O1 to O187), and organized 162 DNA-based O serogroups (O-genotypes) on the basis of the *wzx/wzy* and *wzm/wzt* sequences. Subsequently we presented a comprehensive molecular O-typing scheme: an *E. coli* O-genotyping polymerase chain reaction (ECOG-PCR) system using 20 multiplex PCR sets containing 162 O-genotype-specific PCR primers (Iguchi et al., 2015b).

The Shiga toxin-producing *E. coli* (STEC) constitute one of the most important groups of food-borne pathogens, as they can cause gastroenteritis that may be complicated by hemorrhagic colitis or hemolytic-uremic syndrome (HUS; Tarr et al., 2005). O157 is a leading STEC O serogroup associated with HUS (Terajima et al., 2014; Heiman et al., 2015) and other STEC O serogroups, including O26, O103, O111, O121 and O145, are also recognized as significant food-borne pathogens worldwide (Johnson et al., 2006). Additionally, unexpected STEC O serogroups have sometimes emerged to cause sporadic cases or outbreaks. For example, STEC O104:H4 was responsible for a large food-borne disease outbreak in Europe in Buchholz et al. (2011). For such various O-serogroups, ECOG-PCR is an accurate and reliable approach for subtyping *E. coli* isolates from patients and contaminated foods (Iguchi et al., 2015b; Ombarak et al., 2016). However, as our previous studies indicated, some of the tested strains were not classified into any of the known O genotypes, suggesting the presence of novel O genotypes (Iguchi et al., 2015b).

Here, we analyzed the O-AGCs from genetically untypeable STEC strains (including strains from patients with diarrhea and

hemorrhagic colitis) by the ECOG-PCR. By comparing sequences we revealed six novel O-genotypes and developed specific-PCRs for each novel O-genotype.

MATERIALS AND METHODS

O Serogrouping/O Genotyping

O serogroup were determined by agglutination tests in microtiter plates using commercially available pooled and single antisera against all recognized *E. coli* O antigens (O1 to O187; SSI Diagnostica, 156 Hillerød, Denmark). O genotypes were determined by ECOG-PCR as described in our previous study (Iguchi et al., 2015b). *Salmonella enterica* O42 (SSI Diagnostica) and *Shigella boydii* type 13 (Denka Seiken Co. Ltd., Japan) single antisera were also used to test for the agglutination reaction.

Source Sequences of Novel O-Genotypes

The O-AGC sequences were determined from six O-genotype untypeable (OgUT) STEC strains, of which four were serologically typeable (O1, O39, O40, and O141) and two others were untypeable (OUT) strains (Table 1). All strains were isolated from human feces (including patients with diarrhea and hemorrhagic colitis) in Japan from 2008 to 2012. The O-AGC sequences flanked by *wcaM* and *hisI* were extracted from draft genome sequences determined using an Illumina MiSeq sequencer (Illumina, San Diego, CA, USA), as previously described (Ogura et al., 2015). Identification and functional annotation of the coding sequences were performed based on the results of homology searches against the public non-redundant protein database using BLASTP. Six O-AGC sequences reported in this paper have been deposited in the GenBank/EMBL/DDBJ database (accession no. LC125927-LC125932).

Sequence Comparisons

The *wzx/wzy* sequences from O-serogroup strains (Iguchi et al., 2015a) and OX-groups reference strains (DebRoy et al., 2016) were used. Additionally, O-AGC sequences from O116 (AB812051; Iguchi et al., 2015a), O1 (GU299791; Li et al., 2010), O39 (AB811616; Iguchi et al., 2015a), O141 (DQ868765; Han et al., 2007), O40 (EU296417; Liu et al., 2008), *S. enterica* O42

TABLE 1 | Shiga toxin-producing *E. coli* (STEC) strains used for sequencing of the O-AGC.

Strain ID	Source (case ^a)	Year	O serogroup	O genotype	Novel O genotype	<i>stx1</i>	<i>stx2</i>	<i>eae</i>	Reference
090823	Human (D)	2009	O1	OgUT	OgN10	+	–	–	Iguchi et al., 2015b
100998	Human (BD)	2010	O39	OgUT	OgN31	–	+	–	Iguchi et al., 2015b
121862	Human (D)	2012	O40	OgUT	OgN1	–	+	–	Iguchi et al., 2015b
102755	Human (BD)	2010	O141	OgUT	OgN8	–	+	–	Iguchi et al., 2015b
OT-11	Human (AC)	2008	OUT	OgUT	OgN9	+	–	+	In this study
EHO-67	Human (D)	2011	OUT	OgUT	OgN12	+	–	–	In this study

^aAC, asymptomatic carrier; D, diarrhea; BD, bloody diarrhea.

(JX975340; Liu et al., 2014), and *S. boydii* type 13 (AY369140; Feng et al., 2004) were also used. Multiple alignments of DNA and amino acid sequences were constructed by using the CLUSTAL W program (Thompson et al., 1994). Phylogenetic trees were constructed by using the neighbor-joining algorithm using MEGA5 software (Tamura et al., 2007). Homology comparisons of paired sequences were performed by using the In Silico Molecular Cloning Genomics Edition (In Silico Biology, Inc., Yokohama, Japan).

PCR for Identifying Novel O-Genotypes

Polymerase chain reaction primers for specifically identifying novel O genotypes were designed (Table 2) and their specificities were evaluated by using 185 O-serogroup reference strains (O1–O188) from SSI using the following PCR conditions. Genomic DNA from *E. coli* strains was purified using the Wizard Genomic DNA purification kit (Promega, Madison, WI, USA) or DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. PCRs were performed using 10 ng/μl of template DNA. PCR was performed as follows: each 30-μl reaction mixture contained 2 μl of genomic DNA, 6 μl of 5× Kapa *Taq* buffer, dNTP mix (final concentration, 0.3 mM each), MgCl₂ (final concentration, 2.5 mM), primers (final concentration, 0.5 μM each), and 0.8 U of Kapa *Taq* DNA polymerase (Kapa Biosystems, Woburn, MA, USA). The thermocycling conditions were: 25 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 1 min. PCR products (2 μl) were electrophoresed in 1.5% agarose gels in 0.5× TBE (25 mM Tris borate, 0.5 mM EDTA), and photographed under UV light after the gel was stained with ethidium bromide (1 mg/ml).

Distributional Survey of Novel O-Genotypes

Thirty five O-serogrouped *E. coli* strains from our previous study (Iguchi et al., 2015b), whose O genotypes were not identified by ECOG-PCR were used for screening novel O-genotypes by the PCR method designed in this study. The prevalence of *stx1*, *stx2* (Cebula et al., 1995) and *eae* (Oswald et al., 2000) genes in the tested STEC strains was determined by the PCR.

RESULTS

Novel O-Genotypes

Six types of O-AGC were identified from OgUT STEC strains (Figure 1). Four O-AGCs (named OgN10, OgN31, OgN1, and OgN8 genotypes) were obtained from strains that were serologically classified into O1, O39, O40, and O141, respectively (Table 1). Two others (named OgN9 and OgN12 genotypes) were obtained from strains that were both serologically and genetically unclassified into any groups (Table 1). Actually, the OgN9 strain did not react with any particular antiserum, and OgN12 showed identical agglutination titers with O34 and O140 antisera, which resulted in OUT classification. OgN8, OgN10, and OgN12 carried *rmlBDAC* for the synthesis of deoxythymidine diphosphate (dTDP)-L-rhamnose, and OgN31 carried *rmlBA-vioA* for dTDP viosamine synthesis (Figure 1). OgN9 carried *fmlA-qlnBC* for UDP-N-acetyl-L-quinovosamine (UDP-L-QuiNAc) synthesis (Figure 1). All novel O-AGCs carried the *wzx/wzy* O-antigen processing genes (Figure 1). The *wzx/wzy* sequences from OgN O-AGCs were compared with those from 171 O-serogroup strains and 11 OX-group reference strains,

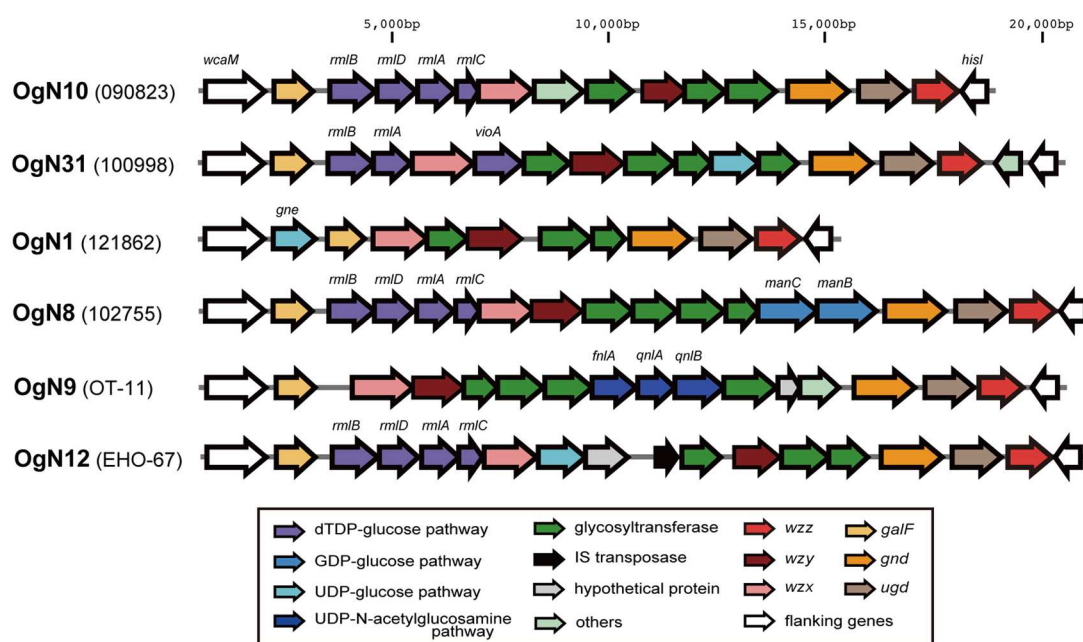


FIGURE 1 | Six novel O-AGCs from STEC strains. The O-AGCs flanked by *wcaM* and *hisI* were extracted from draft genome sequences.

indicating that their sequences were unique compared to those from known O-AGCs (less than 70% DNA sequence identity of closest pairs), except for *wzx* of OgN31 (**Figure 2**). The sequence of OgN31_ *wzx* was 98.7% identical in DNA sequence (99.0% amino acid sequence identity) to that of O116. Sequence comparison of O-AGCs revealed that the left region including

wzx and genes for the d-TDP glucose pathway was conserved between OgN31 and Og116, and the right region including the *wzy* and glycosyltransferase genes was unique (less than 40% DNA sequence identity) in each O-AGC (**Figure 3A**). O-AGC gene sets from four pairs with members of different genotypes that agglutinated with the same O antisera were compared (**Figure 3B**). Between OgN10 and Og1 (from O1 strain), and between OgN8 and Og141 (from O141 strain), *rmlBDAC* genes were highly conserved in both O-AGCs, while other genes including *wzx* and *wzy* were diversified (less than 70% DNA sequence identity). Between OgN31 and Og39 (from O39 strain), different types of sugar biosynthesis genes were located on each O-AGC (*rmlBA-vioA* on OgN31, and *rmlBDAC-vioAB* and *manCB* on Og39). There was no genetic similarity between OgN1 and Og40 (from O40 strain). From these results, we were convinced that these six were novel O-AGCs.

A BLAST search of the NCBI database revealed that the OgN10 O-AGC is similar to that of *S. enterica* O42, and the OgN9 O-AGC was almost identical to that of *S. boydii* type 13 (**Figure 3C**). OgN10 and OgN9 strains agglutinated with *S. enterica* O42 and *S. boydii* type 13 antisera, respectively (data not shown).

Primers for Identifying Novel O Genotypes

Six PCR primer pairs were designed for identifying the novel O-genotypes (**Table 2** and **Figure 4**). All primer pairs were targeted unique sequences of *wzy*, except for OgN31 for which primers were targeted to a glycosyltransferase gene. Each PCR was evaluated by using all 185 O-serogroup reference strains from O1 to O188 and six novel O-genotype strains (listed in **Table 1**). PCR products of the expected sizes on the agarose gel were obtained only with the corresponding strains, and no extra products were observed in the size range between 100 and 1,500 bp (data not shown).

Distribution of Novel O-Genotypes

Among 35 O-serogrouped *E. coli* strains whose O genotypes were not identified by the ECOG-PCR, five O141, three O1, three O39, one O40, and one O140 strains were classified by using the novel O-genotype PCR into OgN8, OgN10, OgN31, OgN1, and OgN12, respectively (**Table 3**). All 13 strains classified into five novel O-genotypes were *eae*-negative STEC, and OgN8, OgN10, and OgN31 had been isolated from patients with bloody diarrhea. OgN8 and OgN12 strains also cross-reacted with O41 and O34 antisera, respectively.

DISCUSSION

In this study, six novel O genotypes were revealed from STEC strains isolated from human patients, and the prevalence of these O genotype strains was confirmed in STECs. The OgN10 strains were serologically classified into the O1 serogroup. The O1-serogroup strain is often seen in extra-intestinal pathogenic *E. coli* from patients with urinary tract infections (Abe et al., 2008; Mora et al., 2009) and septicemic disease (Mora et al.,

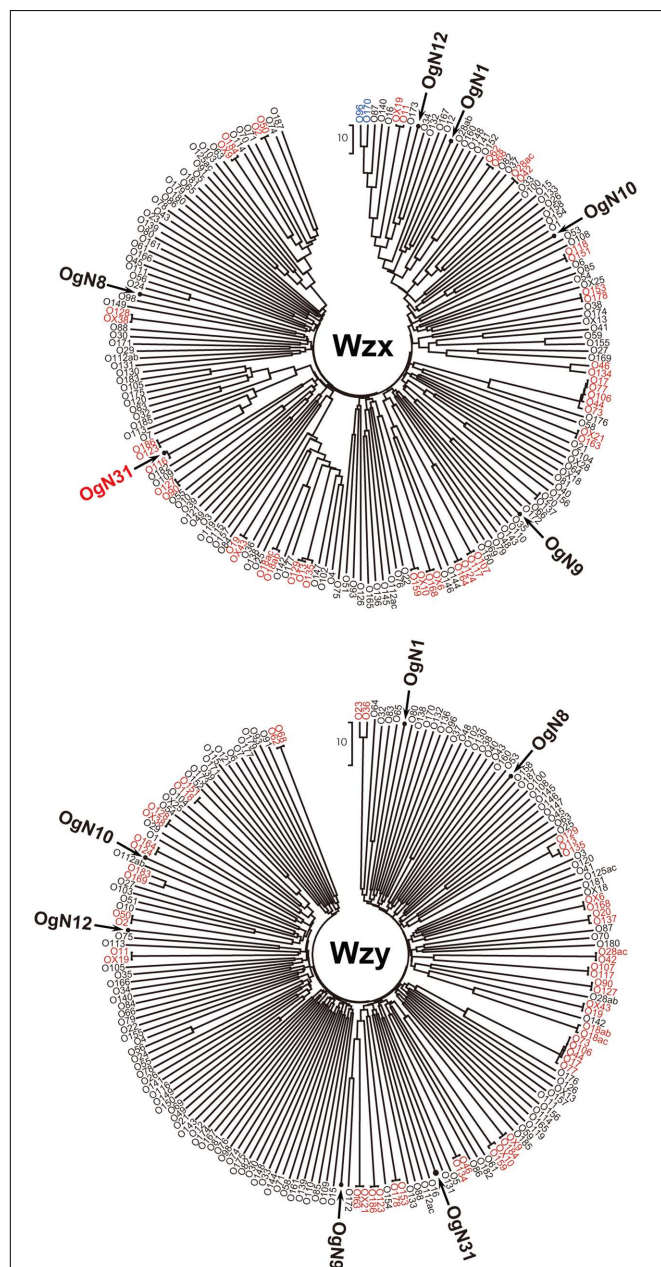


FIGURE 2 | Phylogenetic analysis of Wzx and Wzy homologs from six novel O-AGCs, *Escherichia coli* O-serogroup strains and OX-groups reference strains based on the amino acid sequences (except for *wzy* of OX25, which is not found in the O-AGC of OX25). Pairs or groups of homologs with $\geq 95\%$ DNA sequence identity are indicated in red, and $\geq 70\%$ identity is indicated in blue. Other homologs with low sequence homologies (less than 70%) are indicated in black.

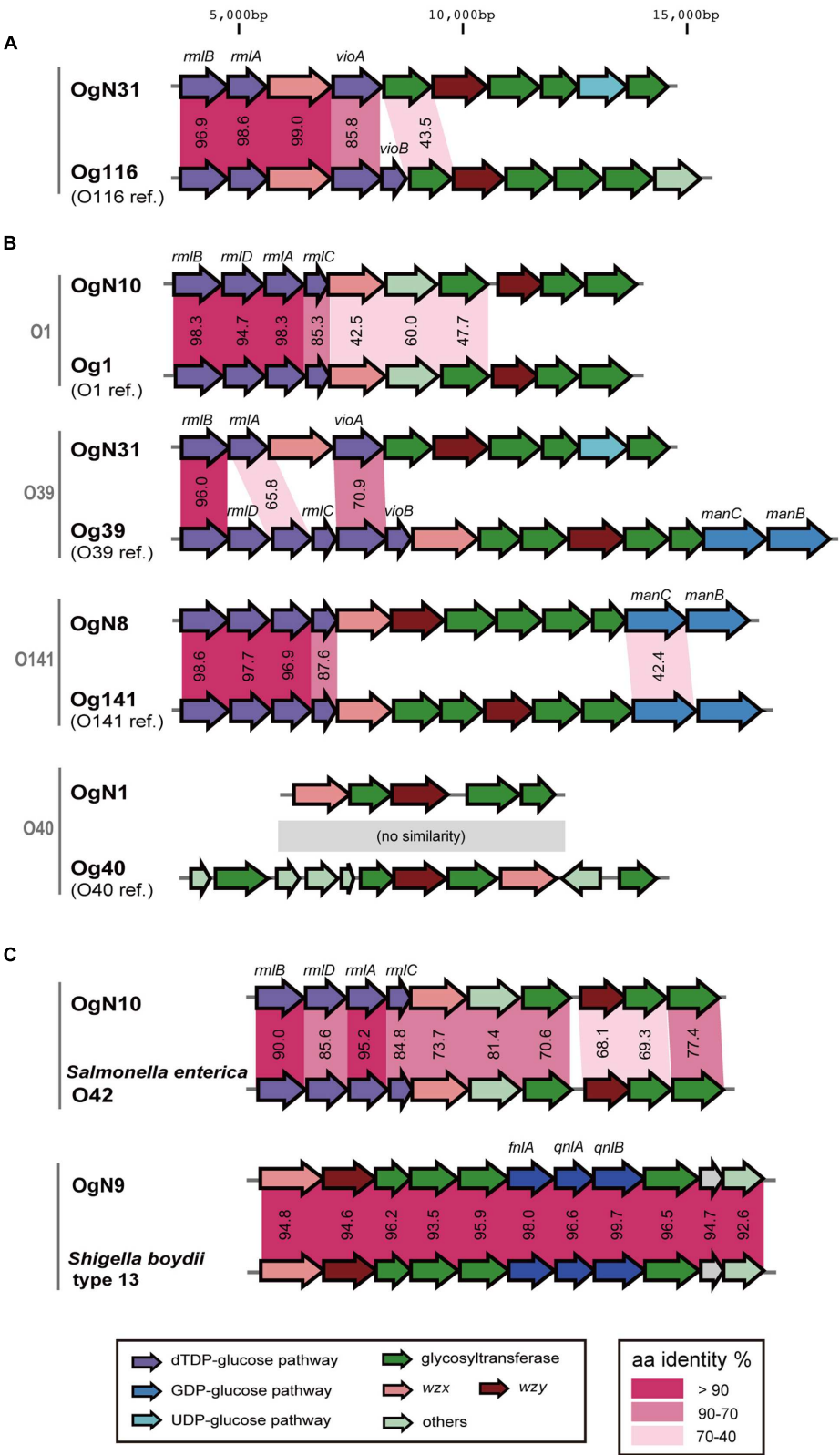
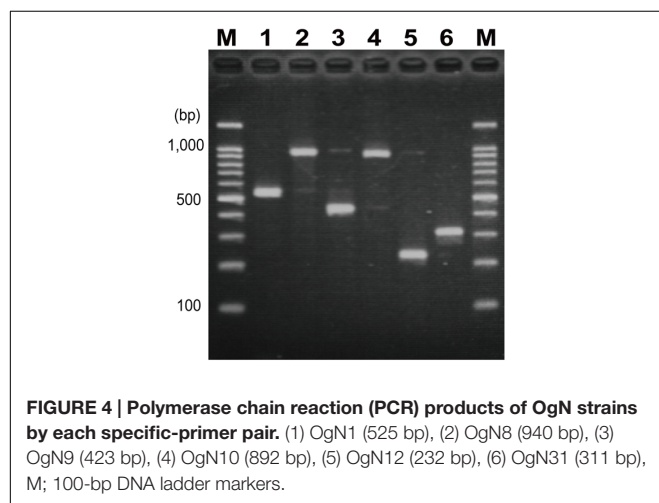


FIGURE 3 | Comparison of O-AGCs. (A) The pair of O-AGCs, OgN31 and O116. **(B)** Four pairs that were serologically agglutinated with the same O antisera carried different types of O-AGCs. Lower genes show the O-AGCs from O-serogroup strains. **(C)** Similar O-AGCs in strains of other genera. Amino acid sequence identities (%) between homologs are shown in the middle.

TABLE 2 | Polymerase chain reaction (PCR) primer sequences for identification of six novel O genotypes.

O genotype	Primer name	Target gene	Sequence (5'–3')	Size (bp)
OgN1	OgN1_PCR_F	wzy	GGTTCCTGTTGCCAATGGT	525
	OgN1_PCR_R		GAGACGAACGTGCAGAAACCA	
OgN8	OgN8_PCR_F	wzy	AACCTTCGCTATGATGGGGG	940
	OgN8_PCR_R		CTTTACCAGGGATGCTCCGA	
OgN9	OgN9_PCR_F	wzy	AAGGTTGGTAGCGTAGGGGA	423
	OgN9_PCR_R		CTCGTATTTCGCCCCCATTC	
OgN10	OgN10_PCR_F	wzy	TGGTGCTGTGTGCTACCATTT	892
	OgN10_PCR_R		AAAGCCAGCCTTAAATCGGA	
OgN12	OgN12_PCR_F	wzy	TTGTGGCACCTGATCCTGCT	223
	OgN12_PCR_R		GCACATGCTAACCCCTGCTCTT	
OgN31	OgN31_PCR_F	Glycosyltransferase	GCCATAAAAAGAGCAAGGGGG	311
	OgN31_PCR_R		GGGGCAGCTGAAAACCAATC	



2009), and in avian pathogenic *E. coli* (Mora et al., 2009; Johnson et al., 2012). Our previous study showed that an O1 strain isolated from patient blood was classified into Og1 (Iguchi et al.,

2015b) and sequence comparisons showed that both an APEC O1 strain from avian colibacillosis (Johnson et al., 2007) and the G1632 strain from a patient with a urinary tract infection (Li et al., 2010) carried the Og1-type O-AGC, whereas three STEC O1 strains used in this study were all classified into OgN10. Actually, we confirmed that five STEC O1 strains from cattle used in a previous study, described as O1B type (Mekata et al., 2014) were also classified into OgN10 (data not shown). In fact, *E. coli* O1 strains could be generally subtyped into two genotypes, Og1 and OgN10, which were clearly linked to extra-intestinal/avian pathogenic *E. coli* and STEC, respectively. Among the O1 serogroup, three types of antigen structures have so far been reported (Baumann et al., 1991; Gupta et al., 1992) and the β -linked side-chain *N*-acetyl-D-mannosamine residue was suggested to be a common O1-specific epitope (Gupta et al., 1992). A partial kinship of O antigen structure synthesized from different O-AGCs may be serologically recognized as the same O serogroup and may also be represented in OgN1 OgN8, OgN12, and OgN31 strains, related to O40, O141, O140, and O39, respectively. The subdivision within each O-serogroup based on the O-AGC DNA sequences may be useful for

TABLE 3 | List of STEC strains classified into novel O genotypes by PCR screening.

Strain ID	Source (case ^a)	Year	O serogroup	O-genotype	Novel O genotype ^b	stx1	stx2	eae
121862	Human (D)	2012	O40	OgUT	OgN1*	–	+	–
102755	Human (BD)	2010	O141	OgUT	OgN8*	–	+	–
110906	Human (AC)	2011	O141	OgUT	OgN8	–	+	–
133345	Human (AC)	2013	O141	OgUT	OgN8	+	+	–
PV10-60	Food	2010	O141	OgUT	OgN8	–	+	–
PV12-70	Food	2012	O141	OgUT	OgN8	–	+	–
072583	Human (BD)	2007	O1	OgUT	OgN10	–	+	–
090823	Human (D)	2009	O1	OgUT	OgN10*	+	–	–
091971	Human (AC)	2009	O1	OgUT	OgN10	+	–	–
091275	Human (AC)	2009	O140	OgUT	OgN12	+	–	–
100998	Human (BD)	2010	O39	OgUT	OgN31*	–	+	–
100999	Human (D)	2010	O39	OgUT	OgN31	–	+	–
NBK#585	Human (AC)	–	O39	OgUT	OgN31	–	+	–

^aAC, asymptomatic carrier; D, diarrhea; BD, bloody diarrhea. ^bStrains used for sequencing of the O-AGC are indicated by asterisks. All strains were isolated in Japan.

obtaining more reliable information for epidemiological studies of pathogenic *E. coli*. Another advantage for DNA-based typing is that serologically untypeable and ambiguous strains could be clearly classified. At the present time, the PCR-based method reported here is the only way to distinguish OgN9. Although, STEC OgN groups have not emerged as a major public health issue, these groups are believed to be a possible cause of diarrhea and bloody diarrhea. To gain more information about trends in STEC OgNs epidemiology, further studies of global OgN isolates are needed. The PCR method described in this study may help the surveillance and monitoring of the OgN groups. Additionally, published sequences from OgN O-AGCs may be used for other DNA-based methodologies, such as *in silico* typing using whole genome sequencing data (Joensen et al., 2015).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: AI, MO, and TH. Performed the experiments: AI, SI, KS, and HN. Analyzed the

data: AI and YO. Contributed reagents/materials/analysis tools: SI, KS, MO, and HM. Wrote the paper: AI. Critical revision of the paper for important intellectual content: AI, SI, and TH.

FUNDING

This research was partially supported by the Research Program on Emerging and Re-emerging Infectious Diseases from Japan Agency for Medical Research and development (AMED; 15fk0108008h0001), and by the Grants-in-Aid for Scientific Research (C) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to AI (25350180) and SI (15K08486).

ACKNOWLEDGMENT

We thank Atsuko Akiyoshi and Yuiko Kato for technical assistance.

REFERENCES

- Abe, C. M., Salvador, F. A., Falsetti, I. N., Vieira, M. A., Blanco, J., Blanco, J. E., et al. (2008). Uropathogenic *Escherichia coli* (UPEC) strains may carry virulence properties of diarrhoeagenic *E. coli*. *FEMS Immunol. Med. Microbiol.* 52, 397–406. doi: 10.1111/j.1574-695X.2008.00388.x
- Baumann, H., Jansson, P. E., Kenne, L., and Widmalm, G. (1991). Structural studies of the *Escherichia coli* O1A O-polysaccharide, using the computer program CASPER. *Carbohydr. Res.* 211, 183–190. doi: 10.1016/0008-6215(91)84159-C
- Beutin, L., Jahn, S., and Fach, P. (2009). Evaluation of the 'GeneDisc' real-time PCR system for detection of enterohaemorrhagic *Escherichia coli* (EHEC) O26, O103, O111, O145 and O157 strains according to their virulence markers and their O- and H-antigen-associated genes. *J. Appl. Microbiol.* 106, 1122–1132. doi: 10.1111/j.1365-2672.2008.04076.x
- Buchholz, U., Bernard, H., Werber, D., Böhmer, M. M., Remschmidt, C., Wilking, H., et al. (2011). German outbreak of *Escherichia coli* O104:H4 associated with sprouts. *N. Engl. J. Med.* 365, 1763–1770. doi: 10.1056/NEJMoa1106482
- Bugarel, M., Beutin, L., Martin, A., Gill, A., and Fach, P. (2010). Microarray for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with Hemorrhagic Colitis and Hemolytic Uremic Syndrome in humans. *Int. J. Food Microbiol.* 142, 318–329. doi: 10.1016/j.ijfoodmicro.2010.07.010
- Cebula, T. A., Payne, W. L., and Feng, P. (1995). Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin type by mismatch amplification mutation assay-multiplex PCR. *J. Clin. Microbiol.* 33, 248–250.
- Coimbra, R. S., Grimont, F., Lenormand, P., Burguière, P., Beutin, L., and Grimont, P. A. (2000). Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). *Res. Microbiol.* 151, 639–654. doi: 10.1016/S0923-2508(00)00134-0
- DebRoy, C., Fratamico, P. M., Yan, X., Baranzoni, G., Liu, Y., Needleman, D. S., et al. (2016). Comparison of O-Antigen Gene Clusters of All O-Serogroups of *Escherichia coli* and Proposal for Adopting a New Nomenclature for O-Typing. *PLoS ONE* 11:e0147434. doi: 10.1371/journal.pone.0147434
- DebRoy, C., Roberts, E., and Fratamico, P. M. (2011a). Detection of O antigens in *Escherichia coli*. *Anim. Health Res. Rev.* 12, 169–185. doi: 10.1017/S1466252311000193
- DebRoy, C., Roberts, E., Valadez, A. M., Dudley, E. G., and Cutter, C. N. (2011b). Detection of Shiga toxin-producing *Escherichia coli* O26, O45, O103, O111, O113, O121, O145, and O157 serogroups by multiplex polymerase chain reaction of the *wzx* gene of the O-antigen gene cluster. *Foodborne Pathog. Dis.* 8, 651–652. doi: 10.1089/fpd.2010.0769
- Feng, L., Senchenkova, S. N., Yang, J., Shashkov, A. S., Tao, J., Guo, H., et al. (2004). Structural and genetic characterization of the *Shigella boydii* type 13 O antigen. *J. Bacteriol.* 186, 383–392. doi: 10.1128/JB.186.2.383-392.2004
- Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., an der Heiden, M., et al. (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* 365, 1771–1780. doi: 10.1056/NEJMoa1106483
- Fratamico, P. M., and Bagi, L. K. (2012). Detection of Shiga toxin-producing *Escherichia coli* in ground beef using the GeneDisc real-time PCR system. *Front. Cell Infect. Microbiol.* 2:152. doi: 10.3389/fcimb.2012.00152
- Geue, L., Monecke, S., Engelmann, I., Braun, S., Slickers, P., and Ehrlich, R. (2014). Rapid microarray-based DNA genosero-typing of *Escherichia coli*. *Microbiol. Immunol.* 58, 77–86. doi: 10.1111/1348-0421.12120
- Gupta, D. S., Shashkov, A. S., Jann, B., and Jann, K. (1992). Structures of the O1B and O1C lipopolysaccharide antigens of *Escherichia coli*. *J. Bacteriol.* 174, 7963–7970.
- Han, W., Liu, B., Cao, B., Beutin, L., Krüger, U., Liu, H., et al. (2007). DNA microarray-based identification of serogroups and virulence gene patterns of *Escherichia coli* isolates associated with porcine postweaning diarrhea and edema disease. *Appl. Environ. Microbiol.* 73, 4082–4088. doi: 10.1128/AEM.01820-06
- Heiman, K. E., Mody, R. K., Johnson, S. D., Griffin, P. M., and Gould, L. H. (2015). *Escherichia coli* O157 Outbreaks in the United States, 2003–2012. *Emerg. Infect. Dis.* 21, 1293–1301. doi: 10.3201/eid2108.141364
- Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., et al. (2015a). A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res.* 22, 101–107. doi: 10.1093/dnares/dsu043
- Iguchi, A., Iyoda, S., Seto, K., Morita-Ishihara, T., Scheutz, F., Ohnishi, M., et al. (2015b). *Escherichia coli* O-Genotyping PCR: a comprehensive and practical platform for molecular o serogrouping. *J. Clin. Microbiol.* 53, 2427–2432. doi: 10.1128/JCM.00321-15
- Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Johnson, K. E., Thorpe, C. M., and Sears, C. L. (2006). The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. *Clin. Infect. Dis.* 43, 1587–1595. doi: 10.1086/509573
- Johnson, T. J., Kariyawasam, S., Wannemuehler, Y., Mangiamale, P., Johnson, S. J., Doetkott, C., et al. (2007). The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J. Bacteriol.* 189, 3228–3236. doi: 10.1128/JB.01726-06

- Johnson, T. J., Wannemuehler, Y., Kariyawasam, S., Johnson, J. R., Logue, C. M., and Nolan, L. K. (2012). Prevalence of avian-pathogenic *Escherichia coli* strain O1 genomic islands among extraintestinal and commensal *E. coli* isolates. *J. Bacteriol.* 194, 2846–2853. doi: 10.1128/JB.06375-11
- Li, D., Liu, B., Chen, M., Guo, D., Guo, X., Liu, F., et al. (2010). A multiplex PCR method to detect 14 *Escherichia coli* serogroups associated with urinary tract infections. *J. Microbiol. Methods* 82, 71–77. doi: 10.1016/j.mimet.2010.04.008
- Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Reeves, P. R., et al. (2014). Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol. Rev.* 38, 56–89. doi: 10.1111/1574-6976.12034
- Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., et al. (2008). Structure and genetics of *Shigella* O antigens. *FEMS Microbiol. Rev.* 32, 627–653. doi: 10.1111/j.1574-6976.2008.00114.x
- Luna-Gierke, R. E., Griffin, P. M., Gould, L. H., Herman, K., Bopp, C. A., Strockbine, N., et al. (2014). Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. *Epidemiol. Infect.* 142, 2270–2280. doi: 10.1017/S0950268813003233
- Mekata, H., Iguchi, A., Kawano, K., Kirino, Y., Kobayashi, I., and Misawa, N. (2014). Identification of O serotypes, genotypes, and virulotypes of Shiga toxin-producing *Escherichia coli* isolates, including non-O157 from beef cattle in Japan. *J. Food Prot.* 77, 1269–1274. doi: 10.4315/0362-028X.JFP-13-506
- Mora, A., López, C., Dabhi, G., Blanco, M., Blanco, J. E., Alonso, M. P., et al. (2009). Extraintestinal pathogenic *Escherichia coli* O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. *BMC Microbiol.* 9:132. doi: 10.1186/1471-2180-9-132
- Ogura, Y., Mondal, S. I., Islam, M. R., Mako, T., Arisawa, K., Katsura, K., et al. (2015). The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Sci. Rep.* 5:16663. doi: 10.1038/srep16663
- Ombarak, R. A., Hinenoya, A., Awasthi, S. P., Iguchi, A., Shima, A., Elbagory, A. R., et al. (2016). Prevalence and pathogenic potential of *Escherichia coli* isolates from raw milk and raw milk cheese in Egypt. *Int. J. Food Microbiol.* 221, 69–76. doi: 10.1016/j.ijfoodmicro.2016.01.009
- Orskov, F., and Orskov, I. (1984). Serotyping of *Escherichia coli*. *Methods Microbiol.* 14, 43–112. doi: 10.1016/S0580-9517(08)70447-1
- Orskov, F., and Orskov, I. (1992). *Escherichia coli* serotyping and disease in man and animals. *Can. J. Microbiol.* 38, 699–704. doi: 10.1139/m92-115
- Oswald, E., Schmidt, H., Morabito, S., Karch, H., Marchès, O., and Caprioli, A. (2000). Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. *Infect. Immun.* 68, 64–71. doi: 10.1128/IAI.68.1.64-71.2000
- Quiñones, B., Swimley, M. S., Narm, K. E., Patel, R. N., Cooley, M. B., and Mandrell, R. E. (2012). O-antigen and virulence profiling of shiga toxin-producing *Escherichia coli* by a rapid and cost-effective DNA microarray colorimetric method. *Front. Cell Infect. Microbiol.* 2:61. doi: 10.3389/fcimb.2012.00061
- Scheutz, F., Cheasty, T., Woodward, D., and Smith, H. R. (2004). Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new *E. coli* O groups that include Verocytotoxin-producing *E. coli* (VTEC): O176, O177, O178, O179, O180 and O181. *APMIS* 112, 569–584. doi: 10.1111/j.1600-0463.2004.apm1120903.x
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Tarr, P. I., Gordon, C. A., and Chandler, W. L. (2005). Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 365, 1073–1086. doi: 10.1016/S0140-6736(05)71144-2
- Terajima, J., Iyoda, S., Ohnishi, M., and Watanabe, H. (2014). Shiga Toxin (Verotoxin)-Producing *Escherichia coli* in Japan. *Microbiol. Spectr.* 2, 1–9. doi: 10.1128/microbiolspec.EHEC-0011-2013
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Wang, Q., Ruan, X., Wei, D., Hu, Z., Wu, L., Yu, T., et al. (2010). Development of a serogroup-specific multiplex PCR assay to detect a set of *Escherichia coli* serogroups based on the identification of their O-antigen gene clusters. *Mol. Cell. Probes* 24, 286–290. doi: 10.1016/j.mcp.2010.06.002
- Wang, S., Meng, Q., Dai, J., Han, X., Han, Y., Ding, C., et al. (2014). Development of an allele-specific PCR assay for simultaneous sero-typing of avian pathogenic *Escherichia coli* predominant O1, O2, O18 and O78 strains. *PLoS ONE* 9:e96904. doi: 10.1371/journal.pone.0096904

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

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

Copyright © 2016 Iguchi, Iyoda, Seto, Nishii, Ohnishi, Mekata, Ogura and Hayashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Characterization of Shiga Toxin Subtypes and Virulence Genes in Porcine Shiga Toxin-Producing *Escherichia coli*

Gian Marco Baranzoni¹, Pina M. Fratamico^{1*}, Jayanthi Gangiredla², Isha Patel², Lori K. Bagi¹, Sabine Delannoy³, Patrick Fach³, Federica Boccia⁴, Aniello Anastasio⁴ and Tiziana Pepe⁴

¹ Eastern Regional Research Center, United States Department of Agriculture – Agricultural Research Service, Wyndmoor, PA, USA, ² Center of Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, MD, USA, ³ Food Safety Laboratory, University of Paris-Est, Anses, Maisons-Alfort, France, ⁴ Department of Veterinary Medicine and Animal Production, University of Naples Federico II, Naples, Italy

OPEN ACCESS

Edited by:

Dustin Brisson,
University of Pennsylvania, USA

Reviewed by:

Jorge Blanco,
University of Santiago
de Compostela, Spain
Seamus Fanning,
University College Dublin, Ireland

*Correspondence:

Pina M. Fratamico
pina.fratamico@ars.usda.gov

Specialty section:

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

Received: 26 February 2016

Accepted: 07 April 2016

Published: 21 April 2016

Citation:

Baranzoni GM, Fratamico PM, Gangiredla J, Patel I, Bagi LK, Delannoy S, Fach P, Boccia F, Anastasio A and Pepe T (2016) Characterization of Shiga Toxin Subtypes and Virulence Genes in Porcine Shiga Toxin-Producing *Escherichia coli*. *Front. Microbiol.* 7:574. doi: 10.3389/fmicb.2016.00574

Similar to ruminants, swine have been shown to be a reservoir for Shiga toxin-producing *Escherichia coli* (STEC), and pork products have been linked with outbreaks associated with STEC O157 and O111:H-. STEC strains, isolated in a previous study from fecal samples of late-finisher pigs, belonged to a total of 56 serotypes, including O15:H27, O91:H14, and other serogroups previously associated with human illness. The isolates were tested by polymerase chain reaction (PCR) and a high-throughput real-time PCR system to determine the Shiga toxin (Stx) subtype and virulence-associated and putative virulence-associated genes they carried. Select STEC strains were further analyzed using a Minimal Signature *E. coli* Array Strip. As expected, *stx*_{2e} (81%) was the most common Stx variant, followed by *stx*_{1a} (14%), *stx*_{2d} (3%), and *stx*_{1c} (1%). The STEC serogroups that carried *stx*_{2d} were O15:H27, O159:H16 and O159:H-. Similar to *stx*_{2a} and *stx*_{2c}, the *stx*_{2d} variant is associated with development of hemorrhagic colitis and hemolytic uremic syndrome, and reports on the presence of this variant in STEC strains isolated from swine are lacking. Moreover, the genes encoding heat stable toxin (*est*_{1a}) and enteroaggregative *E. coli* heat stable enterotoxin-1 (*astA*) were commonly found in 50 and 44% of isolates, respectively. The hemolysin genes, *hlyA* and *ehxA*, were both detected in 7% of the swine STEC strains. Although the *eae* gene was not found, other genes involved in host cell adhesion, including *lpfA*_{O113} and *paa* were detected in more than 50% of swine STEC strains, and a number of strains also carried *iha*, *lpfA*_{O26}, *lpfA*_{O157}, *fedA*, *orfA*, and *orfB*. The present work provides new insights on the distribution of virulence factors among swine STEC strains and shows that swine may carry Stx1a-, Stx2e-, or Stx2d-producing *E. coli* with virulence gene profiles associated with human infections.

Keywords: *Escherichia coli*, STEC, swine, Shiga toxins variants, virulence genes

INTRODUCTION

Shiga Toxin-producing *Escherichia coli* (STEC) are food-borne pathogens responsible for outbreaks and serious illness including hemorrhagic colitis (HC) and hemolytic uremic syndrome (HUS). STEC O157:H7 is the serotype that has most often been associated with outbreaks and severe forms of diarrhea; however, recently a number of non-O157 STEC serogroups that cause similar illnesses have emerged (Gould et al., 2013). Cattle and other ruminants are important reservoirs of STEC; infection is asymptomatic, and they can carry the pathogens for long periods of time. Similarly, healthy swine may shed STEC, as demonstrated by several studies in which STEC were detected and isolated from swine fecal samples (Tseng et al., 2014b). Many of the investigations focused on serotype O157:H7; however, some studies also tested for non-O157 STEC serogroups and identified serogroups previously associated with human cases of illness (Fratamico et al., 2004; Kaufmann et al., 2006; Tseng et al., 2014b). The possibility that swine can transmit pathogenic STEC to humans is supported by a few outbreaks linked to the consumption of pork products contaminated with STEC O157:H7, O157:NM, and O111:H- (Tseng et al., 2014b).

Shiga toxins (Stx) are divided in two major antigenic forms: Stx1 and Stx2. Variants for Stx1 and Stx2 are grouped in three (Stx1a, Stx1c, Stx1d) and seven (Stx2a, Stx2b, Stx2c, Stx2d, Stx2e, Stx2f, and Stx2g) subtypes, respectively (Scheut et al., 2012). Although Stx1a has been linked to human illness, STEC that produce subtypes Stx2a, Stx2c, and Stx2d are more often associated with the development of HC and HUS (Friedrich et al., 2002; Melton-Celsa, 2014). *In vitro* studies in two different cell lines showed that Stx2a and Stx2d were more potent than Stx2b and Stx2c. These results were also confirmed by experimentation in mice showing a significantly higher potency of Stx2a and Stx2d than Stx1, Stx2b, and Stx2c (Fuller et al., 2011). Stx variants are not homogeneously distributed among the STEC population and certain variants are frequently detected in association with different animals (Martin and Beutin, 2011; Hofer et al., 2012; Fuente et al., 2015). Swine STEC strains commonly produce Stx2e (Fratamico et al., 2004; Meng et al., 2014; Tseng et al., 2015), which may cause edema disease in weaned pigs, often leading to ataxia and death. Stx2e-producing *Escherichia coli*, do not represent a particular threat for humans (Friedrich et al., 2002; Tseng et al., 2014b). Nevertheless, STEC carrying the *stx2e* gene have been isolated from human cases with mild diarrhea (Muniesa et al., 2000; Friedrich et al., 2002; Beutin et al., 2004; Sonntag et al., 2005) and from two patients with HUS (Thomas et al., 1994; Fasel et al., 2014). The severe outcome of the first HUS case was probably due to a co-infection with another STEC strain (Thomas et al., 1994), while the second patient with HUS was described as having a very weak immune system (Fasel et al., 2014). Besides Stx2e, there is a lack of information on the presence of other Stx subtypes in STEC strains isolated from swine.

The production of Stx is necessary to provoke HUS; however, other virulence factors are also important in causing illness. These include genes involved in cell adhesion, proteases, and

toxins, as well as other putative virulence factors. The presence of specific combinations of virulence factors may determine the risk of developing severe symptoms. The *eae* gene, found on the locus of enterocyte effacement (LEE), encodes intimin, which is an adhesin involved in gut colonization. LEE-positive STEC are expected to provoke HUS or HC more frequently than LEE-negative STEC (Ethelberg et al., 2004; Toma et al., 2004; Luna-Gierke et al., 2014). Nevertheless, cases of HUS provoked by LEE-negative STEC have been reported (Karmali et al., 1985; Paton et al., 1999; Bielaszewska et al., 2009), including a large outbreak in 2011 in Europe caused by an enteroaggregative *E. coli* that acquired the *stx2a* gene, and it possessed a combination of virulence genes increasing its virulence (Boisen et al., 2015). This suggests that LEE is not essential in the development of severe symptoms, and other genes involved in adherence may also be important. Many adherence gene candidates, including *eibG*, *lpfA*, *saa*, and *sab* have been identified in STEC (Crohen et al., 2013). Nevertheless, mechanisms for attachment of LEE-negative STEC to the intestinal epithelium have not been studied as extensively as attachment of LEE-positive STEC.

In 2000, one objective of the U.S. Department of Agriculture's Animal and Plant Health Inspection Service National Animal Health Monitoring System (NAHMS) Swine 2000 study was to determine the prevalence of STEC in swine. Fecal samples were from states with the highest production of swine in the U.S. (U.S. Department of Agriculture, 2001). As a result of this work, 219 STEC isolates were recovered and characterized (Fratamico et al., 2004, 2008). Since this work was conducted, the knowledge of the importance of non-O157 STEC in human illness has increased, and there is a need to develop a model for molecular risk assessment associated with STEC. Knowledge of the virulence gene combinations that distinguish highly pathogenic *E. coli* from less virulent strains remains unclear, particularly for LEE-negative STEC (Beutin and Fach, 2014). Additionally, new virulence-associated and putative virulence-associated factors are being identified (Coombes et al., 2008; Brandt et al., 2011; Bugarel et al., 2011). The aim of the present study was to characterize STEC recovered from swine, belonging to a variety of serotypes to determine their Stx subtype and virulence gene profiles to understand their virulence potential.

MATERIALS AND METHODS

Bacterial Strains

Swine STEC strains were isolated and serotyped during the NAHMS swine 2000 study (NAHMS 2000) as described by Fratamico et al. (2004). Briefly, fresh swine feces were recovered from the pen floor of swine operations from the main pork-producing states in U.S. A total of 687 swine fecal samples were enriched using tryptic soy broth (TSB) and screened for the presence of *stx1* and *stx2* by polymerase chain reaction (PCR). Positive samples were plated onto Luria-Bertani agar, and *stx1*- and *stx2*-positive colonies were detected following DNA hybridization and confirmed by PCR. Two hundred and nineteen

STEC strains were serotyped and frozen in TSB with 20% of glycerol. From this collection, 181 STEC strains were used in this study and maintained on tryptic soy agar plates or TSB as working stock cultures.

Besides the NAHMS swine isolates, three STEC O91 strains from our collection were also used for comparison. STEC O91:H14 (strains 2.4111 and 2.4114) were isolated from ground beef while STEC O91:H21 (strain B2F1) was isolated from a case of HUS (Ito et al., 1990).

Identification of Shiga-toxin Subtypes

DNA extraction and PCR assays to identify *stx* subtypes and *stx* partial sequences were performed according to Scheutz et al. (2012) using a ProFlex PCR system (Thermo Fisher, Waltham, MA, USA) with slight modifications. TaqMan Environmental Master Mix 2.0 (Thermo Fisher) was used, and the annealing temperature was raised to 65°C when cross-reaction was observed, as suggested by the authors (Scheutz et al., 2012). Gel electrophoresis was performed using 1.5% UltraPure Agarose (Invitrogen, Carlsbad, CA, USA) gel with 0.5X GelRed (Phenix Research Products, Candler, NC, USA) in 1X Tris-acetate-EDTA buffer at 100 V for 1 h. One microliter of amplified DNA was analyzed by agarose gel electrophoresis and visualized using an AlphaImager gel documentation system (Alpha Innotech, San Leandro, CA, USA).

Polymerase chain reaction products for sequencing were cleaned with Agencourt AMPure XP (Beckman Coulter, Brea, CA, USA), and 1.2 µl were amplified in a reaction consisting of 7 µl of 2.5X buffer, 1 µl of 3.2 µM primer *stx2-F4* or *stx2-R1* (Scheutz et al., 2012), 1 µl of Big Dye Terminator (Applied Biosystems), and 10 µl of nuclease-free water. Thermocycling conditions consisted of 30 cycles of 95°C for 10 s, 55°C for 5 s and 60°C for 4 min. The sequencing reaction products were then purified and sequenced using Agencourt CleanSEQ (Beckman Coulter) and 3730 DNA Analyzer (Applied Biosystems), respectively. The sequences were manually curated using Sequencher v5.2.3 (Gene Code Corporation, Ann Arbor, MI, USA), run in VirulenceFinder 1.5 (Joensen et al., 2014), and blasted against the NCBI database¹. The nucleotide sequences were deposited in the GenBank nucleotide sequence database under the following accession numbers: strain 306, KU682619; strain 308, KU682620; strain 326, KU682621; strain 341, KU682622; strain 360, KU682623, and strain 500, KU682624.

High-throughput Real-time PCR Assay and Testing for Hemolysis

DNA was extracted from the swine isolates using the PrepMan Ultra Sample Preparation Reagent (Thermo Fisher) according to the manufacturer's instructions. The high-throughput real-time PCR (hrPCR) assay was carried out using the BioMark real-time PCR system (Fluidigm, San Francisco, CA, USA), targeting 67 virulence-associated and putative virulence-associated genes, 14 O-group-associated genes (O26, O45, O55, O91, O103, O104, O111, O113, O118, O121, O128, O145, O146, and O157) and 11

H-group-associated genes (H2, H4, H7, H8, H11, H16, H19, H21, H25, H28, and H32). Primers were designed in several studies (Perelle et al., 2003; Fratamico et al., 2008; Bugarel et al., 2010, 2011; Delannoy et al., 2013) and summarized by Tseng et al. (2014a). Reagents for DNA amplification and thermal cycling conditions were previously reported (Tseng et al., 2014a). Swine STEC strains positive to *ehxA* and *hlyA* genes were tested for hemolysis by plating onto SHIBAM agar (Hardy Diagnostic, Santa Maria, CA, USA).

FDA Minimal Signature *E. coli* Array

Swine *Stx2d*-producing *E. coli* and non-*Stx2e* STEC belonging to a serotype associated with human disease were further analyzed using the Minimal Signature *E. coli* Array Strip (FDA-ECID; Affymetrix, Santa Clara, CA, USA). Genomic DNA was isolated and concentrated using the DNeasy Tissue Kit (QIAGEN Inc., Valencia, CA, USA) and SC100 Speedvac Concentrator (Savant Instruments, Inc. Holbrook, NY, USA), respectively. Two micrograms of DNA were tested using the FDA-ECID array as described in detail by Lacher et al. (2014). Robust multi-array average summarized probe intensity data were analyzed using R-Bioconductor software v3.1.2 and affy package with parameters defined by Lacher et al. (2014). The Hierarchical clustering was done using overview function in MADE4 package that uses average linkage cluster analysis with a correlation metric distance (Culhane et al., 2005; Culhane and Thioulouse, 2006).

RESULTS

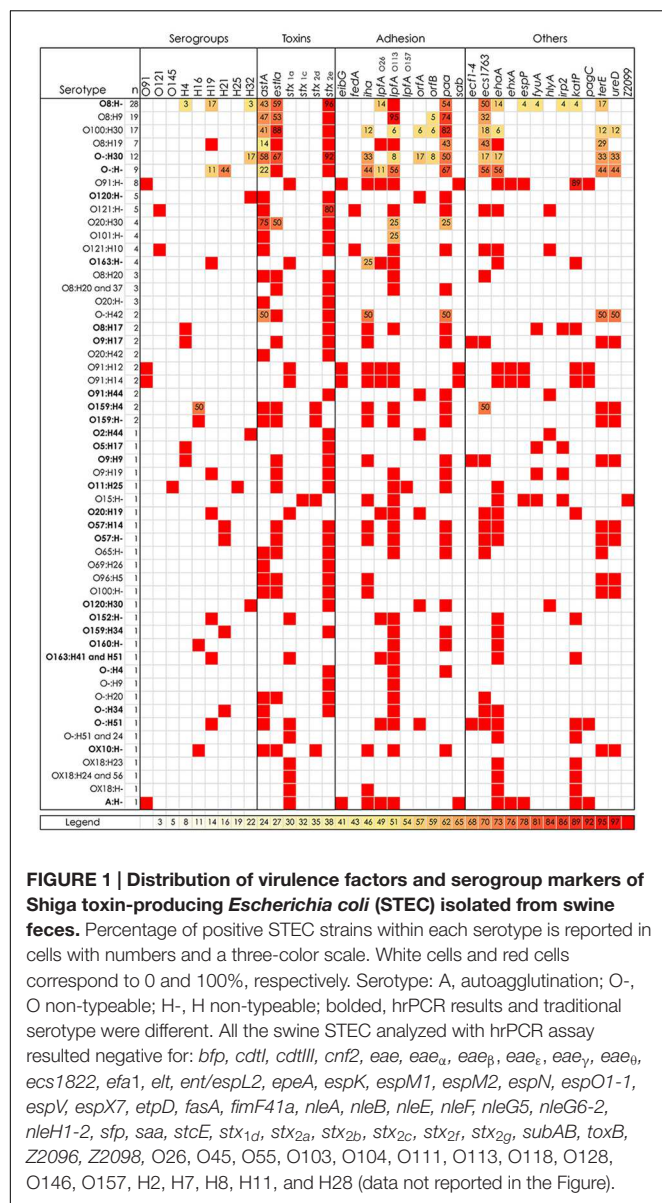
Swine STEC Serotypes

All of the strains had been previously serotyped at the *E. coli* Reference Center at the Pennsylvania State University (University Park, PA, USA). In addition, many O-group- and H-group-specific targets were included in the hrPCR assay. Several discrepancies were found and serotypes that did not match with the traditional serotyping are indicated in bold in **Figure 1**. Selected swine STEC strains were also analyzed using the FDA-ECID microarray, and the resulting serotypes were in agreement with the hrPCR. Moreover, the grouping within the phylogenetic tree was consistent with the serotypes proposed by the FDA-ECID microarray (**Table 1**).

Shiga-toxin Subtype Characterization

The swine STEC strains were analyzed by singleplex and multiplex PCR assays to determine their Shiga-toxin subtype. *Stx*-encoding genes were carried by 177/181 (99.8%) of the tested isolates. Four strains previously identified as STEC likely lost the *Stx* genes due to loss of *Stx*-encoding phages, as has been shown by other investigators (Joris et al., 2011) since PCR results were negative for any of the subtypes. *stx1* or *stx2* genes were carried by 25 and 151 strains, respectively. *Stx* subtype analysis revealed that the 25 *stx1*-positive strains carried the *stx1a* subtype. Among the 151 *stx2*-positive strains, 146 and 5 isolates carried *stx2e* and *stx2d*, respectively. STEC strain 308 was the only isolate that carried both *stx1* and *stx2*, subtypes *stx1c*

¹<http://blast.ncbi.nlm.nih.gov/Blast.cgi>



and *stx2d*, respectively. Stx subtypes divided by STEC serotype are reported in **Figure 1**. Strains carrying Stx subtypes *stx1d*, *stx2a*, *stx2b*, *stx2c*, *stx2f*, and *stx2g* were not identified. Selected swine strains were analyzed using the FDA-ECID microarray and results of Stx subtypes are reported in **Table 1**. Nucleotide sequencing of *stx2* was carried out from STEC strains that were *stx2d* positive by PCR. The STEC strain 308 *stx2* sequence showed 100% identity to a portion of *stx2d* subunit B (AF479829) using VirulenceFinder. When blasted against the NCBI database, the STEC strain 308 *stx2* partial sequence matched EF441621 (*stx2d*) with 100% identity and no gaps. The *stx2* partial sequences of STEC strains 306, 326, 341, 360, and 500 were identical. VirulenceFinder results showed 100% identity with a portion of *stx2d*, subunit B (DQ059012). The most similar sequence present in the NCBI database was KC339670 (*stx2e*) with an identity on 99%.

Distribution of Virulence-associated Genes among the Swine STEC Collection

Genomic DNA extracted from the swine STEC strains was analyzed using hrPCR. Genes encoding the enteroaggregative *E. coli* heat-stable enterotoxin 1 (*astA*) and the heat-stable enterotoxin (*estIa*) were detected in 79 (44%) and 91 (50%) of the isolates, respectively. Toxins and cytotoxic factors encoded by *cdtI*, *cdtIII*, *elt*, *ent/espL2*, *cnf2*, and *subAB* were not detected. Regarding cytotoxins, enterohemolysin (*ehxA*) and α -hemolysin (*hlyA*) encoding genes were found non-simultaneously in 13 (7%) of the swine STEC strains each. These strains were also hemolytic when plated onto SHIBAM plates.

None of the isolated swine STEC strains carried the intimin-encoding gene, *eae*, effector genes involved in the type III secretion system (*espK*, *espM1*, *espM2*, *espN*, *espO1-1*, *espV*, *espX7*, *nleA*, *nleB*, *nleE*, *nleF*, *nleG5*, *nleG6-2*, and *nleH1-2*) or the type II secretion system effector (*etpD*). Other genes encoding factors involved in adhesion and colonization to the host intestine were also investigated. Among these, the most prevalent were *lpfA*_{O113}, *paa*, *ihA*, and *lpfA*_{O26} present in 116 (64%), 98 (54%), 41 (23%), and 33 (18%) isolates, respectively. Genes *orfA*, *orfB*, and *fedA* were detected in less than 8% of the isolates. While *bfp*, *efa1*, *fasA*, *fimF41a*, *saa*, and *toxB* were not found, one swine STEC strain carried *lpfA*_{O157}. Three autotransporter protein genes *ehaA*, *espP*, and *sab* were found in 60 (33%) 13 (7%), and 15 (8%) isolates.

Other gene targets were also included in the high-throughput real-time PCR assay. Positive results were obtained for the *ecs1763*, *terE*, *katP*, and *ureD* genes in 56 (31%), 31 (17%), 27 (15%), and 23 (13%) of isolates, respectively. While less than 8% were positive for *pagC*, *eibG*, *irp2*, *fyuA*, *ecf1*, *ecf2*, *ecf3*, *ecf4*, and Z2099. None of the swine STEC strains carried *ecs1822*, *epeA*, *sfp*, *stcE*, Z2096, or Z2098.

DISCUSSION

It is well-known that swine shed a variety of STEC serogroups, which may be carried along the food production chain. Most of the STEC isolated from these animals have adapted to the swine host and seem to have low potential to infect humans. Nevertheless, outbreaks associated with pork products have occurred (Meng et al., 2014; Tseng et al., 2014b). The sampling area covered by the NAHMS swine 2000 study was large, covering all the main pork-producing States (Fratamico et al., 2004). A subset of 181 STEC strains were analyzed and their pathogenic potential was assessed by detection of virulence and putative virulence factors.

The *stx* subtypes carried by the swine STEC were identified, and the majority of the isolates carried *stx2e* (81%), which was consistent with the data reported by Fratamico et al. (2004). The second most prevalent subtype was *stx1a* (14%), followed by *stx2d* (3%), and *stx1c* (1%). Stx2d is a potent toxin, and infection with strains carrying this subtype can lead to severe symptoms such as HC and HUS in humans (Melton-Celsa, 2014). Besides the Stx genes, the thermostable enterotoxin genes, *astA* and/or *estIa*, genes were found in ~71%

TABLE 1 | Serotypes and phylogenetic tree of select swine STEC strains analyzed using the FDA-ECID array.

Height	FDA-ECID serotype	Serotype (traditional serotyping)	Strain	Isolation (State, Date)	stx ₁ subtype	stx ₂ subtype
0.10 0.00						
	O159:H16	O159:H-	360	NC, 11/28/2000	neg	2e
	O159:H16	O159:H-	306	IO, 10/2/2000	neg	2i
	O159:H16	O159:H4	500	IO, 10/2/2000	neg	2i
	O159:H-	O159:H4	326	KS, 2/6/2001	neg	2i
	O159:H16	OX10:H-	341	MN, 9/19/2000	neg	2i
	O-:H38	O91:H44	448	NE, 10/16/2000	neg	2e
	O-:H38	O91:H44	477	NE, 10/16/2000	neg	2e
	O15:H27	O15:H-	308	IO, 10/2/2000	1c	2acd
	O152:H19	O20:H19	339	MI, 9/20/2000	1a	neg
	O-:H19 ^a	O163:H-	318	IO, 10/17/2000	1a	neg
	O-:H19 ^a	O163:H-	313	IO, 2/5/2001	1a	neg
	O-:H19 ^a	O163:H41 and 51	316	IO, 2/5/2001	1a	neg
	O-:H19 ^a	O163:H-	303	IO, 2/5/2001	1a	neg
	O-:H19 ^a	O163:H-	332	IO, 3/20/2001	1a	neg
	O91:H14	O91:H-	300	IO, 9/18/2000	1a	neg
	O91:H14	O91:H-	337	IO, 10/30/2000	1a	neg
	O91:H14	O91:H-	336	IO, 10/30/2000	1a	neg
	O91:H14	O91:H12	311	SD, 10/2/2000	1a	neg
	O91:H14	O91:H12	312	SD, 10/2/2000	1a	neg
	O91:H14	O91:H-	309	SD, 10/2/2000	1a	neg
	O91:H14	O91:H-	424	IO, 10/30/2000	1a	neg
	O91:H14	O91:H-	307	SD, 10/2/2000	1a	neg
	O91:H14	A:H-	319	IO, 10/16/2000	1a	neg
	O91:H14	O91:H-	315	IO, 10/16/2000	1a	neg
	O91:H14	O91:H-	317	NE, 10/16/2000	1a	neg
	O91:H14 ^b	O91:H14	2.4114	NA	1a	neg
	O91:H14	O91:H14	347	IO, 10/16/2000	1a	neg
	O91:H14 ^b	O91:H14	2.4111	NA	1a	neg
	O91:H14	O91:H14	346	WI, 10/16/2000	1a	neg
	O159:H21	O159:H34	387	IO, 10/2/2000	neg	neg
	O91:H21 ^b	O91:H21	B2F1	NA	neg	2acd

A, autoagglutination; O-, O non-typeable; H-, H non-typeable; A, neg, negative results; NA, not available.

^aThe FDA-ECID does not include probes for O163 markers.

^bSwine STEC isolates were analyzed together with two strains of *E. coli* O91:H14 and *E. coli* O91:H21 for comparison.

of the isolates. Thermostable enterotoxins are usually carried by enterotoxigenic *E. coli*, which are the major pathogens responsible for traveler's diarrhea. Twenty-two percent of the swine STEC strains were positive for both genes. The exotoxins HlyA (α -hemolysin) and EhxA (enterohemolysin) produce pores in the cytoplasmic membranes of eukaryotic cells causing their death. Their role in STEC pathogenesis is still not clear; HlyA may increase the virulence of extraintestinal pathogenic *E. coli* and, in the case of EhxA, a correlation between *ehxA*-positive STEC and development of severe symptoms in humans has been observed (Karch and Bielaszewska, 2001; Mainil, 2013). Thirteen isolates carried the *hlyA* gene. Nine of them belonged to serotypes O121:H- or O121:H10, presenting a virulence gene profile typical of strains associated with edema disease in swine due to the presence of *stx*_{2e}, *hlyA* and *fedA* (Tseng et al., 2014b). The *ehxA* gene is commonly found in STEC.

From 40 to 77% of strains collected from patients, food, and cattle carry this gene (Karch and Bielaszewska, 2001; Slanec et al., 2009; Bosilevac and Koohmaraie, 2011; Feng, 2014). Swine isolates appear to carry *ehxA* less frequently (Meng et al., 2014; Tseng et al., 2014a), and this observation is in agreement with our study where only 7% of the isolates was *ehxA* positive.

All of the swine STEC strains were LEE-negative. Although the adhesion mechanisms of LEE-negative STEC are not well characterized, several factors have been described to play an important role in adhesion to the intestinal epithelium. The long polar fimbriae gene *lpfA*_{O113} was identified in STEC O113:H21 (Doughty et al., 2002). These investigators demonstrated that the removal of *lpfA*_{O113} reduces the bacterial capacity to adhere to epithelial cells. Similar *lpfA* genes were found in *E. coli* O157 and O26 (Hayashi et al., 2001; Toma et al.,

2004). Another bacterial adherence-conferring gene is the iron-regulated gene A homolog adhesin *iha*. Similarly to *lpfA*_{O113}, the *iha* gene is commonly found in STEC strains associated with human cases of HUS (Newton et al., 2009; Galli et al., 2010). Nevertheless, non-pathogenic *E. coli* can also carry *lpfA*_{O113} and *iha*, suggesting that the presence of these genes is insufficient to establish an infection (Toma et al., 2004). Over 80% of the strains analyzed in this study carried *lpfA*_{O26}, *lpfA*_{O113}, or *lpfA*_{O157}; while *iha* was found in almost one quarter of swine isolates. *iha*-positive STEC were also described in a longitudinal study of two Midwestern U.S. pork production sites (Tseng et al., 2014a, 2015). On the contrary, none of the swine STEC strains collected in another interesting study in China carried *iha* (Meng et al., 2014). The second most prevalent adhesion factor found in this dataset was the porcine attaching and effacing-associated adhesin, *paa*, which is associated with neonatal post-weaning diarrhea in pigs (An et al., 1999). In addition, a few strains carried *orfA* and *orfB*, which encode for adhesins involved in diffuse adherence (Charbonneau et al., 2006).

Autotransporter proteins have a peculiar structure that allows them to move through the membrane system and execute their function outside the bacterial cell. The genes *ehaA* and *sab* were discovered in O157:H7 strain EDL933 and LEE-negative O113:H21, respectively. They encode for two different autotransporter proteins that contribute to adhesion and biofilm formation (Wells et al., 2008; Herold et al., 2009). Together with LEE genes, *iha* and *ehaA* are highly expressed in the intestines of pigs presenting attaching and effacing lesions (Liu et al., 2015). While the *ehaA* gene was present in over 30% of the swine isolates, *sab* was carried by 13 STEC strains only belonging to O-group O91.

As stated above, 12 to 18% of the isolates were positive for *katB*, *ureD*, and *terE*. The genes *katP* and *ureD* encode for a catalase/peroxidase and urease transporter, respectively. Their role in *E. coli* pathogenesis is unclear; however, they appear to be prevalent in diarrheagenic *E. coli* (Dorothea et al., 2006; Delannoy et al., 2013). The gene *terE* is a component of the *ter* cluster, which confers tellurite resistance (Orth et al., 2007). The *ecs1763* and *ecs1822* genes have been proposed to be novel markers for enterohemorrhagic *E. coli*. Their function is unknown, and they were shown to be shared by a clonal group of enterohemorrhagic *E. coli* that includes O26, O111, and O118 (Abu-Ali et al., 2009). Tseng et al. (2014a) observed that *ecs1763* is frequently found in swine STEC, which was confirmed by the present study where 31% of the isolates carried *ecs1763*. *ecs1822* was absent in all the tested strains.

Traditional serotyping of *E. coli* is time consuming, and cross-reactions among antisera often occur. Based on the hrPCR and FDA-ECID results, 71 strains present in this collection belonged to serotypes O8:H9, O8:H19, O8:H-, O15:H27, O20:H-, O91:H14, O101:H-, O121:H-, O145:H25, O159:H21, and O163:H19 that were previously isolated from human patients (Blanco et al., 1992; Beutin and Fach, 2014). Serotypes O8:H19, O15:H27, O145:H25, and O163:H19 have also been associated with cases of HUS (Prager et al., 2005; Bielaszewska et al., 2006; Galli et al., 2010). All of the strains belonging to the

serotypes O8:H9, O8:H19, O8:H-, O20:H-, O101:H-, O121:H-, O145:H25, and O159:H21 analyzed in this study carried *stx*_{2e}, which is a subtype that is generally not associated with STEC that cause serious human illness. Human infections linked to Stx2e-producing *E. coli* generally cause asymptomatic infections or mild diarrhea (Tseng et al., 2014b). The work of Sonntag et al. (2005) reported that human Stx2e-producing *E. coli* carry different virulence factors compared to swine Stx2e-producing *E. coli* associated with edema disease. They also detected *fyuA* and *irp2* genes in five strains isolated from humans. These genes are included in the high-pathogenicity island (HPI), which is involved in the iron metabolism of *Yersinia*. Mouse models showed that the HPI increases *E. coli* virulence in extraintestinal infections (Schubert et al., 2002). The hrPCR results revealed that some swine STEC strains belonged to the same serotypes as human Stx2e-producing *E. coli* (O8:H19 and O8:H-) reported by Sonntag et al. (2005). STEC O8:H19 and STEC O8:H- also carried markers for the HPI. Moreover, their virulence gene profiles included adhesins (*lpfA*_{O26}, *lpfA*_{O113}, *paa*) and enterotoxins (*astA* and *estIa*), which suggest that they can potentially provoke mild diarrhea in humans. The HPI genes *fyuA* and *irp2* were also found in Stx2e-producing *E. coli* belonging to serotypes O5:H4 and O8:H4 (Table 1).

Shiga toxin-producing *Escherichia coli* strain 308 was re-typed as O15:H27 using the FDA-ECID array and was found to have the same *stx*_{2d} sequence as *E. coli* O15:H27 (strain 88-1509) in the STEC isolate database at Michigan State University². *E. coli* strain 88-1509 was collected in 1988 from a human case of HC and HUS in Canada. Other strains belonging to serotype O15:H27 have been isolated from human and cattle feces, and from meat sources (Piérard et al., 1997; Woodward et al., 2002; Bosilevac et al., 2007; Galli et al., 2010). The LEE-negative swine STEC O15:H27 has a virulence gene profile consisting of *stx*_{1c}, *stx*_{2d}, *ehaA*, *espP*, *fyuA*, *iha*, *irp2*, *lpfA*_{O113}, and Z2099. The relevance of some of these genes was mentioned above. *E. coli* secreted protein P (EspP) is an autotransporter protein with serine protease activity, and is used by the bacteria to impair the complement response of the host (Orth et al., 2010). Recently, In et al. (2013) reported that EspP boosts macropinocytosis in the intestinal epithelium increasing Stx uptake. The open reading frame Z2099 is highly prevalent in typical and emerging enterohemorrhagic *E. coli*, while it is significantly less prevalent in non-pathogenic *E. coli* (Delannoy et al., 2013).

Six of the swine STEC strains carried *stx*_{2d} according to PCR and VirulenceFinder results, and they belonged to serotypes O159:H-, O159:H4, and OX10:H-. DebRoy et al. (2016) reported that serological cross-reactions between the O159 and OX10 O-groups often occur and that the nucleotide sequences of O159 and OX10 O-antigen gene clusters are almost identical. Based on the FDA-ECID analysis, the strains 306, 360, 341, and 500 were re-typed as O159:H16; while the strain 326 was re-typed as O159:H- (Table 1). STEC belonging to O-group O159 rarely infect humans (Brooks et al., 2005; Gould et al., 2013). STEC O159:H16 and O159:H- have been isolated only

²<http://shigatox.net/new/database.html>

from swine samples, such as feces and carcasses (DesRosiers et al., 2001; Kaufmann et al., 2006; Meng et al., 2014). Stx subtype analysis of these strains often gives ambiguous results (Kaufmann et al., 2006; Meng et al., 2014). In this work, STEC O159:H16 and O159:H- were positive for *stx_{2d}* when tested by PCR; however, they were positive for *stx_{2e}* or *stx_{2i}* using the FDA-ECID array. Note that probes of the FDA-ECID array corresponding to *stx_{2i}* were designed using the Stx sequences AM904726 and FN252457 (Patel et al., 2016) that belong to the *stx_{2e}* subtype according to Scheut et al. (2012). The product obtained from partial sequencing of *stx₂* was 99% identical to the sequence KC339670 when blasted against the NCBI database. KC339670 is a complete *stx₂* sequence belonging to a STEC O159:H16 strain isolated from swine in China. After a neighbor-joining cluster analysis of the sequence, Meng et al. (2014) concluded that KC339670 represented a new variant of *stx_{2e}*. Further investigations using cell lines and animal models are needed to understand the virulence potential of this Stx2 variant. Another STEC O159 was detected in this collection. It belonged to the H21 H-group and was positioned distantly from the clade of O159:H16 and O159:H- (Table 1). This strain was positive for *stx_{2e}* only by PCR. *E. coli* belonging to serotype O159:H21 was isolated in 1983 during a small outbreak of diarrhea involving newborn children in Spain (Blanco et al., 1992), and no other infections associated with serotype O159:H21 have been reported.

Locus of enterocyte effacement-negative STEC belonging to O-group O91 are frequently associated with adult human infections with symptoms ranging from mild diarrhea to HC and HUS. The main serotypes are O91:H14 and O91:H21, and the latter is usually linked with development of severe symptoms (Bielaszewska et al., 2009). Human STEC O91:H14 and O91:H21 isolates carried mainly *stx₁* and *stx_{2d}*, respectively (Prager et al., 2005; Bielaszewska et al., 2009; Galli et al., 2010). These STEC have been isolated from food samples derived from bovine, swine, and ovine origin, and from both domestic and wild animals (Martin and Beutin, 2011; Ju et al., 2012). From the NAHMS swine 2000 study, 15 strains belonging to serotypes O91:H12, O91:H14, O91:H44, and O91:H- were isolated from fresh fecal samples collected from four different states (Fratamico et al., 2004). Eight of these strains were re-typed as O91:H14, while STEC O91:H44 strains 448 and 477 did not belong to the O91 O-group based on FDA-ECID and hrPCR results (Figure 1; Table 1). STEC strain 319 that carried an identical virulence gene profile to other O91:H14 strains was also re-typed as O91:H14 by FDA-ECID array (Table 1). According to the phylogenetic tree in Table 1, the clade of STEC O91:H14 strains is well separated from the STEC O91:H21 strain B2F1 isolated from a human case of HUS. Interestingly, two O91:H14 strains were more closely related to two STEC O91:H14 strains isolated from ground beef samples than the other swine STEC O91:H14 strains. Despite the fact that one strain was *katP*-negative, all 13 STEC O91:H14 strains presented a conserved virulence gene profile (*ehaA*, *ehxA*, *eibG*, *espP*, *ihaA*, *katP*, *lpfA_{O26}*, *lpfA_{O113}*, *pagC*, *sab*, and *stx_{1a}*), which is very similar to profiles of strains from human clinical samples (Prager et al., 2005; Bielaszewska et al., 2009). Similar to *ihaA*, *lpfA_{O26}*, and *lpfA_{O113}*, the proteins encoded by the genes

eibG and *sab* are involved in host gut colonization. The *E. coli* immunoglobulin-binding protein encoded by *eibG* binds human immunoglobulin G and immunoglobulin A, and contributes to epithelial host cell adhesion (Lu et al., 2006), and *sab* is a gene encoding for an autotransporter protein involved in biofilm formation and found in a pathogenic LEE-negative STEC (Herold et al., 2009). Lastly, the *pagC* gene encodes for an outer membrane protein present in different Enterobacteriaceae that contributes to serum resistance (Nishio et al., 2005).

Human infections caused by STEC O163:H19 are rare (Brooks et al., 2005; Gould et al., 2013). However, Stx2-producing *E. coli* O163:H19 provoked sporadic cases of HUS (Prager et al., 2005) and have been found associated with cattle and produce (Woodward et al., 2002; Galli et al., 2010; Feng, 2014). In this work, five strains of STEC O163:H- or O163:H41/H51 were re-typed as O163:H19. They all carried *stx_{1a}* similar to the Stx1-producing *E. coli* O163:H19 strain isolated from swine by DesRosiers et al. (2001).

STEC O20:H19 is associated with human cases of HUS (Galli et al., 2010), and one strain belonging to this serotype was isolated in the NAHMS study (Fratamico et al., 2004). However, this same strain was re-analyzed using the FDA-ECID array, and it was re-typed as O152:H19, which is not known to be a human pathogen.

CONCLUSION

Using state-of-the-art DNA-based techniques, this study provides new insights on the distribution of virulence factors in a heterogeneous collection of STEC isolated from the major pork-producing states of the U.S. Stx2e-producing *E. coli* known to provoke mild diarrhea in humans carried different virulence factors than Stx2e-producing *E. coli* associated with edema disease in pigs; this finding suggests that Stx2e-producing *E. coli* that cause human illnesses may not have a swine origin (Sonntag et al., 2005). In our work, STEC strains carrying *stx_{2e}* belonging to the same serotype and having similar virulence gene profiles as Stx2e-producing *E. coli* isolated from humans were identified. Additionally, the majority of Stx2e-producing *E. coli* carried thermostable enterotoxin genes usually found in enterotoxigenic *E. coli*.

This work suggests that STEC, including serotypes O15:H27 and O91:H14 that have been associated with human illness and are found in multiple hosts or environments, could also be carried by swine. Interestingly, a strain of O15:H27 found to carry *stx_{2d}* and other virulence genes may have the potential to produce severe symptoms in humans. Moreover, STEC O91:H14 strains presented a virulence gene profile very similar to profiles found in human isolates.

AUTHOR CONTRIBUTIONS

GMB and PF. designed research; GMB, LKB, SD, PF, FB, AA, and TP performed research; GMB, JG, and IP analyzed data; GMB and PF wrote the paper.

ACKNOWLEDGMENTS

This research was supported in part by an appointment to the Agricultural Research services (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the

USDA. ORISE is managed by ORAU under DOE contract number DE-AC05-06OR23100. The hrPCR development was partially financed by the French joint ministerial program of R&D against CBRNE risks. All opinions expressed in this manuscript are the authors' and do not necessarily reflect the policies and views of USDA, ARS, DOE, or ORAU/ORISE.

REFERENCES

- Abu-Ali, G. S., Lacher, D. W., Wick, L. M., Qi, W., and Whittam, T. S. (2009). Genomic diversity of pathogenic *Escherichia coli* of the EHEC 2 clonal complex. *BMC Genom.* 10:296. doi: 10.1186/1471-2164-10-296
- An, H., Fairbrother, J. M., Desautels, C., and Harel, J. (1999). "Distribution of a novel locus called Paa (porcine attaching and effacing associated) among Enteric *Escherichia coli*," in *Mechanisms in the Pathogenesis of Enteric Diseases 2 Advances in Experimental Medicine and Biology*, eds P. S. Paul and D. H. Francis (Boston, MA: Springer), 179–184. doi: 10.1007/978-1-4615-4143-1
- Beutin, L., and Fach, P. (2014). Detection of Shiga toxin-producing *Escherichia coli* from nonhuman sources and strain typing. *Microbiol. Spectr.* 2:EHEC-0001-2013. doi: 10.1128/microbiolspec.EHEC-0001-2013
- Beutin, L., Krause, G., Zimmermann, S., Kaulfuss, S., and Gleier, K. (2004). Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. *J. Clin. Microbiol.* 42, 1099–1108. doi: 10.1128/JCM.42.3.1099
- Bielaszewska, M., Friedrich, A. W., Aldick, T., Schürk-Bulgrin, R., and Karch, H. (2006). Shiga toxin activatable by intestinal mucus in *Escherichia coli* isolated from humans: predictor for a severe clinical outcome. *Clin. Infect. Dis.* 43, 1160–1167. doi: 10.1086/508195
- Bielaszewska, M., Stoewe, F., Fruth, A., Zhang, W., Prager, R., Brockmeyer, J., et al. (2009). Shiga toxin, cytolethal distending toxin, and hemolysin repertoires in clinical *Escherichia coli* O91 isolates. *J. Clin. Microbiol.* 47, 2061–2066. doi: 10.1128/JCM.00201-09
- Blanco, J., González, E. A., Espinosa, P., Blanco, M., Garabal, J. I., and Alonso, M. P. (1992). Enterotoxigenic and necrotizing *Escherichia coli* in human diarrhoea in Spain. *Eur. J. Epidemiol.* 8, 548–552. doi: 10.1007/BF00146375
- Boisen, N., Melton-Celsa, A. R., Scheutz, F., O'Brien, A. D., and Nataro, J. P. (2015). Shiga toxin 2a and enteroaggregative *Escherichia coli* – A deadly combination. *Gut Microbes* 6, 272–278. doi: 10.1080/19490976.2015.1054591
- Bosilevac, J. M., Guerini, M. N., Brichta-Harhay, D. M., Arthur, T. M., and Koohmaraie, M. (2007). Microbiological characterization of imported and domestic boneless beef trim used for ground beef. *J. Food Prot.* 70, 440–449.
- Bosilevac, J. M., and Koohmaraie, M. (2011). Prevalence and characterization of non-O157 shiga toxin-producing *Escherichia coli* isolates from commercial ground beef in the United States. *Appl. Environ. Microbiol.* 77, 2103–2112. doi: 10.1128/AEM.02833-10
- Brandt, S. M., King, N., Cornelius, A. J., Premaratne, A., Besser, T. E., and On, S. L. W. (2011). Molecular risk assessment and epidemiological typing of Shiga toxin-producing *Escherichia coli* by using a novel PCR binary typing system. *Appl. Environ. Microbiol.* 77, 2458–2470. doi: 10.1128/AEM.02322-10
- Brooks, J. T., Sowers, E. G., Wells, J. G., Greene, K. D., Griffin, P. M., Hoekstra, R. M., et al. (2005). Non-O157 Shiga toxin – producing *Escherichia coli* infections in the United States, 1983 – 2002. *J. Infect. Dis.* 192, 1422–1424. doi: 10.1086/466536
- Bugarel, M., Beutin, L., and Fach, P. (2010). Low-density microarray targeting non-locus of enterocyte effacement effectors (nle genes) and major virulence factors of Shiga toxin-producing *Escherichia coli* (STEC): a new approach for molecular risk assessment of STEC isolates. *Appl. Environ. Microbiol.* 76, 203–211. doi: 10.1128/AEM.01921-09
- Bugarel, M., Martin, A., Fach, P., and Beutin, L. (2011). Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis for molecular risk assessment of typical and atypical EPEC strains. *BMC Microbiol.* 11:142. doi: 10.1186/1471-2180-11-142
- Charbonneau, M. È., Berthiaume, F., and Mourez, M. (2006). Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I). *J. Bacteriol.* 188, 8504–8512. doi: 10.1128/JB.00864-06
- Coombes, B. K., Wickham, M. E., Mascarenhas, M., Gruenheid, S., Finlay, B. B., and Karmali, M. A. (2008). Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing *Escherichia coli* strains. *Appl. Environ. Microbiol.* 74, 2153–2160. doi: 10.1128/AEM.02566-07
- Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., and Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* 26, 822–880. doi: 10.1128/CMR.00022-13
- Culhane, A. C., and Thioulouse, J. (2006). A multivariate approach to integrating datasets using made4 and ade4. *News. R Proj.* 6, 54–58.
- Culhane, A. C., Thioulouse, J., Perrière, G., and Higgins, D. G. (2005). MADE4: An R package for multivariate analysis of gene expression data. *Bioinformatics* 21, 2789–2790. doi: 10.1093/bioinformatics/bti394
- DeRoy, C., Fraticamo, P. M., Yan, X., Baranzoni, G., Liu, Y., Needleman, D. S., et al. (2016). Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-typing. *PLoS ONE* 11:e0147434. doi: 10.1371/journal.pone.0147434
- Delannoy, S., Beutin, L., and Fach, P. (2013). Discrimination of enterohemorrhagic *Escherichia coli* (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. *J. Clin. Microbiol.* 51, 3257–3262. doi: 10.1128/JCM.01471-13
- DesRosiers, A., Fairbrother, J. M., Johnson, R. P., Desautels, C., Letellier, A., and Quessy, S. (2001). Phenotypic and genotypic characterization of *Escherichia coli* verotoxin-producing isolates from humans and pigs. *J. Food Prot.* 64, 1904–1911. doi: 10.1128/JCM.01095-08
- Dorothea, O., Grif, K., Dierich, M. P., and Wu, R. (2006). Prevalence, structure and expression of urease genes in Shiga toxin-producing *Escherichia coli* from humans and the environment. *Int. J. Hyg. Environ. Health* 209, 513–520. doi: 10.1016/j.ijheh.2006.06.003
- Doughty, S., Sloan, J., Bennett-Wood, V., Robertson, M., Robins-Browne, R. M., and Hartland, E. L. (2002). Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic *Escherichia coli*. *Infect. Immun.* 70, 6761–6769. doi: 10.1128/IAI.70.12.6761-6769.2002
- Ethelberg, S., Olsen, K. E. P., Scheutz, F., Jensen, C., Schiellerup, P., Engberg, J., et al. (2004). Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg. Infect. Dis.* 10, 842–847. doi: 10.3201/eid1005.030576
- Fasel, D., Mellmann, A., Cernela, N., Hächler, H., Fruth, A., Khanna, N., et al. (2014). Hemolytic uremic syndrome in a 65 year-old male linked to a very unusual type of stx2e and eae harboring O51:H49 Shiga-toxin producing *Escherichia coli*. *J. Clin. Microbiol.* 52, 1301–1303. doi: 10.1128/JCM.03459-13
- Feng, P. (2014). Shiga Toxin-Producing *Escherichia coli* (STEC) in fresh produce — A food safety dilemma. *Microbiol. Spectr.* 2:EHEC-0010-2013. doi: 10.1128/microbiolspec.EHEC-0010
- Fraticamo, P. M., Bagi, L. K., Bush, E. J., and Solow, B. T. (2004). Prevalence and characterization of Shiga toxin-producing *Escherichia coli* in swine feces recovered in the National Animal Health Monitoring system's swine 2000 study. *Appl. Environ. Microbiol.* 70, 7173–7178. doi: 10.1128/AEM.70.12.7173-7178.2004
- Fraticamo, P. M., Bhagwat, A. A., Injaian, L., and Fedorka-Cray, P. J. (2008). Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from swine feces. *Foodborne Pathog. Dis.* 5, 827–838. doi: 10.1089/fpd.2008.0147
- Friedrich, A. W., Bielaszewska, M., Zhang, W., Pulz, M., Kuczius, T., Ammon, A., et al. (2002). *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J. Infect. Dis.* 185, 74–84. doi: 10.1086/338115

- Fuente, R. D. La Horcajo, P., and Dom, G. (2015). Association of vt1c with verotoxin-producing *Escherichia coli* from goats and sheep. *J. Vet. Diagn. Invest.* 334, 332–334.
- Fuller, C. A., Pellino, C. A., Flagler, M. J., Strasser, J. E., and Weiss, A. A. (2011). Shiga toxin subtypes display dramatic differences in potency. *Infect. Immun.* 79, 1329–1337. doi: 10.1128/IAI.01182-10
- Galli, L., Miliwebsky, E., Irino, K., Leotta, G., and Rivas, M. (2010). Virulence profile comparison between LEE-negative Shiga toxin-producing *Escherichia coli* (STEC) strains isolated from cattle and humans. *Vet. Microbiol.* 143, 307–313. doi: 10.1016/j.vetmic.2009.11.028
- Gould, L. H., Mody, R. K., Ong, K. L., Clogher, P., Cronquist, A. B., Garman, K. N., et al. (2013). Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000–2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Pathog. Dis.* 10, 453–460. doi: 10.1089/fpd.2012.1401
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., et al. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22. doi: 10.1093/dnares/8.1.47
- Herold, S., Paton, J. C., and Paton, A. W. (2009). Sab, a novel autotransporter of locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* O113:H21, contributes to adherence and biofilm formation. *Infect. Immun.* 77, 3234–3243. doi: 10.1128/IAI.00031-09
- Hofer, E., Cernela, N., and Stephan, R. (2012). Shiga toxin subtypes associated with shiga toxin-producing *Escherichia coli* strains isolated from red deer, roe deer, chamois, and ibex. *Foodborne Pathog. Dis.* 9, 792–795. doi: 10.1089/fpd.2012.1156
- In, J., Lukyanenko, V., Foulke-Abel, J., Hubbard, A. L., Delannoy, M., Hansen, A.-M., et al. (2013). Serine protease EspP from enterohemorrhagic *Escherichia coli* is sufficient to induce Shiga toxin macropinocytosis in intestinal epithelium. *PLoS ONE* 8:e69196. doi: 10.1371/journal.pone.0069196
- Ito, H., Terai, A., Kurazono, H., Takeda, Y., and Nishibuchi, M. (1990). Cloning and nucleotide sequencing of Vero toxin 2 variant genes from *Escherichia coli* O91:H21 isolated from a patient with the hemolytic uremic syndrome. *Microb. Pathog.* 8, 47–60. doi: 10.1016/0882-4010(90)90007-D
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Joris, M. -A., Verstraete, K., De Reu, K., and De Zutter, L. (2011). Loss of vtx genes after the first subcultivation step of verocytotoxigenic *Escherichia coli* O157 and Non-O157 during isolation from naturally contaminated fecal samples. *Toxins (Basel)* 3, 672–677. doi: 10.3390/toxins3060672
- Ju, W., Shen, J., Li, Y., Toro, M. A., Zhao, S., Ayers, S., et al. (2012). Non-O157 Shiga toxin-producing *Escherichia coli* in retail ground beef and pork in the Washington D.C. area. *Food Microbiol.* 32, 371–377. doi: 10.1016/j.fm.2012.07.017
- Karch, H., and Bielaszewska, M. (2001). Sorbitol-fermenting Shiga toxin-producing *Escherichia coli* O157:H(-) strains: epidemiology, phenotypic and molecular characteristics, and microbiological diagnosis. *J. Clin. Microbiol.* 39, 2043–2049. doi: 10.1128/JCM.39.6.2043-2049.2001
- Karmali, M. A., Petric, M., Lim, C., Fleming, P. C., Arbus, G. S., and Lior, H. (1985). The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. *J. Infect. Dis.* 151, 775–782. doi: 10.1093/infdis/151.5.775
- Kaufmann, M., Zweifel, C., Blanco, M., Blanco, J. E., Beutin, L., et al. (2006). *Escherichia coli* O157 and non-O157 Shiga toxin-producing *Escherichia coli* in fecal samples of finished pigs at slaughter in Switzerland. *J. Food Prot.* 69, 260–266.
- Lacher, D. W., Gangiredla, J., Jackson, S. A., Elkins, C. A., and Feng, P. C. H. (2014). Novel microarray design for molecular serotyping of Shiga toxin-producing *Escherichia coli* isolated from fresh produce. *Appl. Environ. Microbiol.* 80, 4677–4682. doi: 10.1128/AEM.01049-14
- Liu, B., Yin, X., Yu, H., Feng, Y., Ying, X., Gong, J., et al. (2015). Alteration of the microbiota and virulence gene expression in *E. coli* O157:H7 in pig ligated intestine with and without AE lesions. *PLoS ONE* 10:e0130272. doi: 10.1371/journal.pone.0130272
- Lu, Y., Iyoda, S., Satou, H., Satou, H., Itoh, K., Saitoh, T., et al. (2006). A new immunoglobulin-binding protein, EibG, is responsible for the chain-like adhesion phenotype of locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. *Infect. Immun.* 74, 5747–5755. doi: 10.1128/IAI.00724-06
- Luna-Gierke, R. E., Griffin, P. M., Gould, L. H., Herman, K., Bopp, C. A., Strockbine, N., et al. (2014). Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. *Epidemiol. Infect.* 142, 2270–2280. doi: 10.1017/S0950268813003233
- Mainil, J. (2013). *Escherichia coli* virulence factors. *Vet. Immunol. Immunopathol.* 152, 2–12. doi: 10.1016/j.vetimm.2012.09.032
- Martin, A., and Beutin, L. (2011). Characteristics of Shiga toxin-producing *Escherichia coli* from meat and milk products of different origins and association with food producing animals as main contamination sources. *Int. J. Food Microbiol.* 146, 99–104. doi: 10.1016/j.ijfoodmicro.2011.01.041
- Melton-Celsa, A. R. (2014). Shiga toxin (Stx) classification, structure, and function. *Microbiol. Spectr.* 2, 1–13. doi: 10.1128/microbiolspec.EHEC-0024-2013
- Meng, Q., Bai, X., Zhao, A., Lan, R., Du, H., Wang, T., et al. (2014). Characterization of Shiga toxin-producing *Escherichia coli* isolated from healthy pigs in China. *BMC Microbiol.* 14:5. doi: 10.1186/1471-2180-14-5
- Muniesa, M., Recktenwald, J., Bielaszewska, M., Karch, H., and Schmidt, H. (2000). Characterization of a Shiga toxin 2e-converting bacteriophage from an *Escherichia coli* strain of human origin. *Infect. Immun.* 68, 4850–4855. doi: 10.1128/IAI.68.9.4850-4855.2000
- Newton, H. J., Sloan, J., Bulach, D. M., Seemann, T., Allison, C. C., Tauschek, M., et al. (2009). Shiga toxin-producing *Escherichia coli* strains negative for locus of enterocyte effacement. *Emerg. Infect. Dis.* 15, 372–380. doi: 10.3201/eid1502.080631
- Nishio, M., Okada, N., Miki, T., Haneda, T., and Danbara, H. (2005). Identification of the outer-membrane protein PagC required for the serum resistance phenotype in *Salmonella enterica* serovar Choleraesuis. *Microbiology* 151, 863–873. doi: 10.1099/mic.0.27654-0
- Orth, D., Ehrlenbach, S., Brockmeyer, J., Khan, A. B., Huber, G., Karch, H., et al. (2010). EspP, a serine protease of enterohemorrhagic *Escherichia coli*, impairs complement activation by cleaving complement factors C3/C3b and C5. *Infect. Immun.* 78, 4294–4301. doi: 10.1128/IAI.00488-10
- Orth, D., Grif, K., Dierich, M. P., and Würzner, R. (2007). Variability in tellurite resistance and the ter gene cluster among Shiga toxin-producing *Escherichia coli* isolated from humans, animals and food. *Res. Microbiol.* 158, 105–111. doi: 10.1016/j.resmic.2006.10.007
- Patel, I. R., Gangiredla, J., Lacher, D. W., Mammel, M. K., Jackson, S. A., Lampel, K. A., et al. (2016). FDA-*Escherichia coli* identification (FDA-ECID) microarray: a pan-genome molecular toolbox for serotyping, virulence profiling, molecular epidemiology, and phylogeny. *Appl. Environ. Microbiol.* (in press). doi: 10.1128/AEM.04077-15
- Paton, A. W., Woodrow, M. C., Doyle, R. M., Lanser, J. A., and Paton, J. C. (1999). Molecular characterization of a Shiga toxigenic *Escherichia coli* O113:H21 strain lacking eae responsible for a cluster of cases of hemolytic-uremic syndrome. *J. Clin. Microbiol.* 37, 3357–3361.
- Perelle, S., Dilasser, F., Grout, J., and Fach, P. (2003). Development of a 5'-nuclease PCR assay for detecting Shiga toxin-producing *Escherichia coli* O145 based on the identification of an "O-island 29" homologue. *J. Appl. Microbiol.* 94, 587–594. doi: 10.1046/j.1365-2672.2003.01872.x
- Piérard, D., Stevens, D., Moriau, L., Lior, H., and Lauwers, S. (1997). Isolation and virulence factors of verocytotoxin-producing *Escherichia coli* in human stool samples. *Clin. Microbiol. Infect.* 3, 531–540. doi: 10.1111/j.1469-0691.1997.tb00303.x
- Prager, R., Annemüller, S., and Tschäpe, H. (2005). Diversity of virulence patterns among Shiga toxin-producing *Escherichia coli* from human clinical cases - need for more detailed diagnostics. *Int. J. Med. Microbiol.* 295, 29–38. doi: 10.1016/j.ijmm.2004.12.009
- Scheutz, F., Teel, L. D., Beutin, L., Piérard, D., Buvens, G., Karch, H., et al. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12

- Schubert, S., Picard, B., Gouriou, S., Heesemann, J., and Denamur, E. (2002). Yersinia high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infect. Immun.* 70, 5335–5337. doi: 10.1128/IAI.70.9.5335
- Slanec, T., Fruth, A., Creuzburg, K., and Schmidt, H. (2009). Molecular analysis of virulence profiles and Shiga toxin genes in food-borne Shiga toxin-producing *Escherichia coli*. *Appl. Environ. Microbiol.* 75, 6187–6197. doi: 10.1128/AEM.00874-09
- Sonntag, A. K., Bielaszewska, M., Mellmann, A., Dierksen, N., Schierack, P., Wieler, L. H., et al. (2005). Shiga toxin 2e-producing *Escherichia coli* isolates from humans and pigs differ in their virulence profiles and interactions with intestinal epithelial cells. *Appl. Environ. Microbiol.* 71, 8855–8863. doi: 10.1128/AEM.71.12.8855-8863.2005
- Thomas, A., Cheasty, T., Chart, H., and Rowe, B. (1994). Isolation of Vero cytotoxin-producing *Escherichia coli* serotypes O9ab:H- and O101:H-carrying VT2 variant gene sequences from a patient with haemolytic uraemic syndrome. *Eur. J. Clin. Microbiol. Infect. Dis.* 13, 1074–1076. doi: 10.1007/BF02111832
- Toma, C., Martínez Espinosa, E., Song, T., Miliwebsky, E., Chinen, I., Iyoda, S., et al. (2004). Distribution of putative adhesins in different seropathotypes of Shiga toxin-producing *Escherichia coli*. *J. Clin. Microbiol.* 42, 4937–4946. doi: 10.1128/JCM.42.11.4937-4946.2004
- Tseng, M., Fratamico, P. M., Bagi, L., Delannoy, S., Fach, P., Manning, S. D., et al. (2014a). Diverse virulence gene content of Shiga toxin-producing *Escherichia coli* from finishing swine. *Appl. Environ. Microbiol.* 80, 6395–6402. doi: 10.1128/AEM.01761-14
- Tseng, M., Fratamico, P. M., Bagi, L., Manzinger, D., and Funk, J. A. (2015). Shiga toxin-producing *E. coli* (STEC) in swine: prevalence over the finishing period and characteristics of the STEC isolates. *Epidemiol. Infect.* 143, 505–514. doi: 10.1017/S0950268814001095
- Tseng, M., Fratamico, P. M., Manning, S. D., and Funk, J. A. (2014b). Shiga toxin-producing *Escherichia coli* in swine: the public health perspective. *Anim. Health Res. Rev.* 15, 1–13. doi: 10.1017/S1466252313000170
- U.S. Department of Agriculture. (2001). Part I. *Reference of Swine Health and Health Management in the United States, 2000*. Fort Collins, CO: National Animal Health Monitoring System. Publication no. 338.0801.
- Wells, T. J., Sherlock, O., Rivas, L., Mahajan, A., Beatson, S. A., Torpdahl, M., et al. (2008). EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157:H7 that contributes to adhesion and biofilm formation. *Environ. Microbiol.* 10, 589–604. doi: 10.1111/j.1462-2920.2007.01479.x
- Woodward, D. L., Clark, C. G., Caldeira, R. A., Ahmed, R., and Rodgers, F. G. (2002). Verotoxigenic *Escherichia coli* (VTEC): A major public health threat in Canada. *Can. J. Infect. Dis.* 13, 321–330. doi: 10.1155/2002/383840

Disclaimer: Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

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

Copyright © 2016 Baranzoni, Fratamico, Gangiredla, Patel, Bagi, Delannoy, Fach, Boccia, Anastasio and Pepe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Amit Kumar,
Kansas State University, USA
T. G. Nagaraja,
Kansas State University, USA

*Correspondence:

Joseph M. Bosilevac,
U. S. Department of Agriculture,
Agricultural Research Service, Roman
L. Hruska U. S. Meat Animal Research
Center, State Spur 18D, Clay Center,
NE 68933-0166, USA
mick.bosilevac@ars.usda.gov

[†]USDA is an equal opportunity
provider and employer. Trade names
are necessary to report factually on
available data; however, the USDA
neither guarantees nor warrants the
standard of the product, and the use
of the name by USDA implies no
approval of the product to the
exclusion of others that may also be
suitable.

Specialty section:

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

Received: 13 July 2015

Accepted: 15 September 2015

Published: 29 September 2015

Citation:

Luedtke BE and Bosilevac JM (2015)
Comparison of methods for the
enumeration of enterohemorrhagic
Escherichia coli from veal hides and
carcasses. *Front. Microbiol.* 6:1062.
doi: 10.3389/fmicb.2015.01062

Comparison of methods for the enumeration of enterohemorrhagic *Escherichia coli* from veal hides and carcasses

Brandon E. Luedtke[†] and Joseph M. Bosilevac^{*†}

U. S. Department of Agriculture, Agricultural Research Service, Roman L. Hruska U. S. Meat Animal Research Center, Clay Center, NE, USA

The increased association of enterohemorrhagic *Escherichia coli* (EHEC) with veal calves has led the United States Department of Agriculture Food Safety and Inspection Service to report results of veal meat contaminated with the Top 7 serogroups separately from beef cattle. However, detection methods that can also provide concentration for determining the prevalence and abundance of EHEC associated with veal are lacking. Here we compared the ability of qPCR and a molecular based most probable number assay (MPN) to detect and enumerate EHEC from veal hides at the abattoir and the resulting pre-intervention carcasses. In addition, digital PCR (dPCR) was used to analyze select samples. The qPCR assay was able to enumerate total EHEC in 32% of the hide samples with a range of approximately 34 to 91,412 CFUs/100 cm² (95% CI 4–113,460 CFUs/100 cm²). Using the MPN assay, total EHEC was enumerable in 48% of the hide samples and ranged from approximately 1 to greater than 17,022 CFUs/100 cm² (95% CI 0.4–72,000 CFUs/100 cm²). The carcass samples had lower amounts of EHEC with a range of approximately 4–275 CFUs/100 cm² (95% CI 3–953 CFUs/100 cm²) from 17% of samples with an enumerable amount of EHEC by qPCR. For the MPN assay, the carcass samples ranged from 0.1 to 1 CFUs/100 cm² (95% CI 0.02–4 CFUs/100 cm²) from 29% of the samples. The correlation coefficient between the qPCR and MPN enumeration methods indicated a moderate relation ($R^2 = 0.39$) for the hide samples while the carcass samples had no relation ($R^2 = 0.002$), which was likely due to most samples having an amount of total EHEC below the reliable limit of quantification for qPCR. Interestingly, after enrichment, 81% of the hide samples and 94% of the carcass samples had a detectable amount of total EHEC by qPCR. From our analysis, the MPN assay provided a higher percentage of enumerable hide and carcass samples, however determining an appropriate dilution range and the limited throughput offer additional challenges.

Keywords: veal, total EHEC, multiplex qPCR, MPN, dPCR

Introduction

Shiga-toxin producing *Escherichia coli* (STEC) are an increasing concern in relation to food safety. The United States Department of Agriculture (USDA) Food Safety and Inspection Service (FSIS) has identified the pathogenic strains of the serogroups O26, O45, O103, O111, O121, and O145 (Top 6) in addition to O157 as being adulterants in non-intact beef (Almanza, 2011). However, emerging STEC serogroups pose a threat to human health with emphasis on the STEC subgroup that comprises the enterohemorrhagic *E. coli* (EHEC). Approximately 20% of human illnesses caused by a non-O157 EHEC were attributed to a serogroup not identified by FSIS (Brooks et al., 2005; Gould, 2009). The EHEC serogroups mostly cause the severest form of disease and can result in hemorrhagic colitis and/or hemolytic uremic syndrome primarily in children under 10 and the elderly (Goldwater and Bettelheim, 2012). In the environment, cattle act as the primary reservoir for EHEC and facilitate the transmission of the bacteria through the release of contaminated feces. Moreover, during the harvesting of cattle, EHEC can contaminate the carcass via the transfer of feces from the animal hide (Elder et al., 2000; Monaghan et al., 2012).

Recently, FSIS has placed interest in the increased association of the adulterant EHEC with veal products compared to beef (United States Department of Agriculture and Food Safety and Inspection Service, 2012) and has implicated a hide to carcass transmission as the primary mode of contamination (United States Department of Agriculture, and Food Safety and Inspection Service, 2013). Indeed, among weaned beef calves entering the feedlot environment the fecal prevalence of O157:H7 was found to be at 5% while 54% of tested hides were positive for O157:H7 (Arthur et al., 2009). However, a study investigating total STEC prevalence found 100% of 62 white veal calves were positive by ELISA for Shiga toxin 1 and/or Shiga toxin 2 (*stx1/2*) (Cristancho et al., 2008). Although isolation and genetic characterization of the *stx1/2* strains was not conducted for these samples, it does suggest that veal calves have the potential to harbor EHEC amongst the total STEC and could lead to hide contamination prevalence greater than that of O157:H7. The limited studies involving non-O157:H7 have identified EHEC of the serogroups O26, O103, O111, O118, and O145 as being associated with calves (Wieler et al., 1998; Pearce et al., 2004; Wang et al., 2014). This is likely not an exhaustive list of EHEC serogroups and additional studies are required to elucidate other EHEC serogroups found in veal calves. In addition, the current method for enumerating EHEC from veal calf samples uses direct plate counts on selective media and is limited to O157:H7, hence molecular assays to detect and enumerate EHEC associated with veal calves are required (Wang et al., 2014).

Culture based enumeration strategies, such as most probable number (MPN) or direct plate counts, can be a subjective and time-consuming process. Moreover, these assays could be impacted by EHEC that are viable but not culturable. Although, the contribution of viable but not culturable, EHEC to human disease is not fully known (Ramamurthy et al., 2014), molecular based assays would detect and include the unculturable EHEC in the enumeration. The use of real time PCR (qPCR) based

enumeration methods are common for samples recovered from cattle. These assays primarily target a combination of the genes *stx1/2*, intimin (*eae*), *uidA*, *rfbE*, and *fliC* alleles for the detection and/or enumeration of O157:H7 and select Top 6 serogroups (Jacob et al., 2012; Wasilenko et al., 2012). However, these genes can be found separately in cells that are non-EHEC. Recently we used the *E. coli* attaching and effacing gene-positive conserved fragment 1 (*ecf1*), which is solely associated with EHEC (Boerlin et al., 1999; Becker and Groschel, 2014), as a gene target for the detection and enumeration of total EHEC directly from cattle feces using qPCR and reported a reliable limit of quantification of 1.25×10^3 CFUs/mL (Luedtke et al., 2014). To provide a lower limit of detection, the *ecf1* target could be utilized in a molecular based modified MPN assay (Russo et al., 2014). In addition, the third generation of PCR termed digital PCR (dPCR) may offer an advantage to qPCR for the enumeration of total EHEC. In dPCR, Poisson based statistics are used to quantify absolute amounts of target DNA from tens of thousands of sub-nanoliter sized endpoint PCR reactions per sample. This reaction partitioning limits the interference of PCR inhibitors, allows for the detection of rare targets, and is not prone to amplification variability like replicate qPCR Cq values in the 35–40 range (Huggett et al., 2013; Marx, 2014).

Here we are the first to utilize three distinct molecular based assays detecting *ecf1* to enumerate total EHEC associated with veal hides at the abattoir and the resulting carcass. For the enumeration of total EHEC, identical pre-enrichment samples were run in parallel using our previously mentioned qPCR assay, a molecular based MPN assay, and by dPCR. Each assay had comparative strengths and weaknesses. The MPN assay provided the highest detection and enumeration rate while the qPCR assay allowed for the greatest dynamic range. Moreover, this is the first reported use of dPCR for the direct enumeration of bacteria from cattle samples. The application of a rapid and accurate assay for the enumeration of total EHEC associated with veal could provide a valuable tool, which is currently lacking.

Materials and Methods

Sample Collection

Paired hide and carcass samples were collected from 95 20- to 22-week-old formula fed veal calves at a veal processing plant in December 2013. Prior to any form of microbial intervention, hide samples were obtained after stunning using Speci-Sponges (Nasco, Fort Atkinson, WI), moistened with buffered peptone water (BPW) (BD, Sparks, MD), to swab an approximate 500 cm² area over the breast-plate region. The sponges were passed (back and forth counting as one pass) five times either vertically or horizontally within the sample area and the sponge was flipped and passed five times in the remaining direction. The sponges were placed in respective Whirl-Pak bags (Nasco) containing 20 mL BPW. After the hide microbial intervention and removal, the respective carcass samples were obtained before any additional intervention in a similar fashion as the hide samples. An approximate 6000 cm² area from the inside and outside round and the navel-plate-brisket-foreshank areas was swabbed and placed in Whirl-Pak bags with 20 mL BPW. All

samples were secured in insulated coolers with ice packs for transport to the United States Meat Animal Research Center. All samples were processed the following day. Before removing aliquots for analysis, bacteria were dislodged from the sponges and suspended by roughly hand massaging the Whirl-Pak bags.

qPCR

Prior to the enrichment of the hide and carcass samples, 20 μ l of the respective sample was added to 180 μ l of the BAX[®] system lysis buffer containing BAX[®] system protease and then prepared using the manufacturer's guidelines (DuPont, Wilmington, DE). All DNA samples were stored at -20°C prior to processing. A standard curve was generated using the *E. coli* O157:H7 reference strain EDL 932 (ATCC 43894) and divided into single use aliquots that were stored at -20°C . The EDL 932 standards, hide and carcass samples, and no template controls were run in duplicate reactions using the duplex qPCR assay targeting *eae* and *ecf1* as previously described (Luedtke et al., 2014). To normalize the quantification of the total EHEC across separate reaction plates, a pooled approach was used to develop the standard curve for enumeration. The pooled approach was reported to reduce uncertainty in the concentration of an unknown sample compared to a standard curve generated from a single instrument run since a similar mean is established from all of the instrument runs in the study analysis (Sivaganesan et al., 2010). The total EHEC was recorded as CFUs/mL and then converted to CFUs/100 cm^2 using the previously described equation (Bohaychuk et al., 2011). In addition, the theoretical reliable limit of enumeration was calculated as CFUs/100 cm^2 of swabbed hide and carcass using the previously described limit of enumeration of 1250 CFUs/mL for the *ecf1* target (Luedtke et al., 2014).

Modified Most Probable Number

A modified most probable number assay (MPN) was developed to increase the sensitivity of detection for the enumeration of total EHEC. From the Whirl-Pak bags, a 1 mL aliquot was transferred to 3 mLs of Tryptic Soy broth (TSB). For the hide samples, this initial dilution was used to create additional triplicate dilutions of 1:44, 1:484, and 1:5324 in BPW and incubated for 6 h at 42°C . Since the carcass samples likely had a lower starting amount of total EHEC, triplicate dilutions of 1:4, 1:44, and 1:484 were created and incubated as previously described. A 1 mL portion of each dilution for the hide and carcass samples was inoculated into a Roka G2 Sample Transfer Tube (Roka Bioscience, San Diego, CA). Samples were shipped on ice to the Roka Bioscience laboratory for analysis using the automated Atlas[®] system (Roka Bioscience), which targets *ecf1* mRNA for subsequent transcription mediated amplification and a hybridization protection assay. The MPN was determined from the number *ecf1* of positive replicates for each dilution and calculated using a freeware MPN calculator (Jarvis et al., 2010). The total EHEC was recorded as CFUs/mL and then converted to CFUs/100 cm^2 using the previously described equation (Bohaychuk et al., 2011). Four carcass samples were incorrectly loaded and were removed from the sample set and comparative analysis.

dPCR

Since this is the first description of using dPCR to enumerate bacteria directly from environmental sources, select samples with enumerable total EHEC from the MPN and qPCR were utilized for absolute enumeration. The absolute enumeration of total EHEC from 26 hide and 16 carcass pre-enrichment samples was performed in 15 μ l reactions. The reactions contained 7.5 μ l of the Quantstudio[™] 3D Digital PCR Master Mix (Applied Biosystems[®] by Life Technologies, Carlsbad, CA), 2 μ l of DNA, 3.7 μ l of PCR grade H_2O , and the addition of the *eae* and *ecf1* primers and probes at the previously described concentrations (Luedtke et al., 2014). A 14.5 μ l aliquot of each reaction was loaded onto respective QuantStudio[™] 3D Digital PCR 20K Chips (Applied Biosystems[®] by Life Technologies), which has 20,000 wells that can accommodate a 865 pL reaction per well, using the automated QuantStudio[™] 3D Digital PCR Chip Loader (Applied Biosystems[®] by Life Technologies). No template controls and a field sample that screened negative by qPCR for both targets were also included for each run. All loaded chips were assembled according to the manufacturer's recommendations (Applied Biosystems[®] by Life Technologies). For the amplification of the DNA targets, the loaded chips were placed on a flat block Gene Amp[®] 9700 thermocycler (Applied Biosystems[®] by Life Technologies) and used the cycling conditions 96°C for 10 min, with 39 cycles of 59°C for 2 min and 98°C for 30 s, a hold at 59°C for 2 min, and a 4°C hold. After the thermocycling was completed, the chips were allowed to warm to room temperature and analyzed within 1 hr after removal from the thermocycler using the QuantStudio[™] 3D Digital PCR Instrument (Applied Biosystems[®] by Life Technologies). Samples where the chip leaked the QuantStudio[™] 12K Flex OpenArray[®] Immersion Fluid (Applied Biosystems[®] by Life Technologies) during thermocycling were redone using a new chip. Further analysis of the data was performed using the Quantstudio[™] 3D AnalysisSuite[™] version 2.0.0 (Applied Biosystems[®] by Life Technologies). All chips were analyzed for the quality of the read and adjusted to a quality threshold of 0.6 for an increased stringency. The fluorescent threshold was adjusted according to the no template control and the sample that screened negative for both targets by qPCR. These controls served as a baseline for background fluorescence of the FAM and MAXN dyes. The fluorescent threshold was adjusted above the background and then universally for *eae* and *ecf1* across all samples due to the mono-modal peak associated with samples that contain a limited amount of target DNA. The total EHEC was recorded as CFUs/mL and then converted to CFUs/100 cm^2 using the previously described equation (Bohaychuk et al., 2011).

Sample Enrichment

To determine the prevalence of total EHEC in the hide and carcass samples, the Whirl-Pak bags were supplemented with 80 mL of TSB and incubated for 6 h at 42°C . The enriched samples were processed for qPCR as previously described and a multiplex qPCR assay was used to detect the presence of *eae*, *ecf1*, and *stx1/2* (Luedtke et al., 2014).

Statistics

GraphPad Prism 6 (GraphPad Software, La Jolla, CA) was used to determine the 95% confidence intervals for the qPCR assays, construct the Bland-Altman plots, Pearson correlation coefficients, and was used to calculate significant differences using the χ^2 test. *P*-values <0.05 were considered significant.

Results

qPCR Standard Curve and Enumeration of total EHEC from Veal Hide and Carcass Samples

Using a pooled approach to develop the qPCR standard curve for enumerating total EHEC provided a reproducible curve over the five log dilution series (Table 1). Overall, the PCR efficiency for *eae* and *ecf1* was 102 and 104%, respectively (Figure 1). In addition, the no template controls were consistently negative across all qPCR assays.

Total EHEC was enumerable in 30 (32%) of the 95 pre-enrichment hide samples using the duplex qPCR assay. Based on the *ecf1* target, the amount of total EHEC ranged from approximately 34 to 91,412 CFUs/100 cm² (95% CI 4–113,460 CFUs/100 cm²). However, 26 (87%) of the enumerable samples were below the calculated reliable limit of enumeration of 5000 CFUs/100 cm² for the hide samples. Despite being below the reliable limit of enumeration, total EHEC could be enumerated based on extrapolation from the standard curve. However, some samples had a single C_q value from the duplicate reactions. This also occurred amongst the carcass samples. From the 95 pre-enrichment carcass samples, total EHEC was enumerable in 16 (17%) samples and all of the enumerable samples had a concentration below the calculated reliable limit of quantification of approximately 417 CFUs/100 cm² for the carcass samples. Using the *ecf1* target, the amount of total EHEC in the carcass samples ranged from approximately 4–275 CFUs/100 cm² (95% CI 3–953 CFUs/100 cm²).

MPN Enumeration of Total EHEC from Veal Hide and Carcass Samples

The MPN assay was able to enumerate total EHEC in 46 (48%) of the 95 hide swab samples, and indicated a concentration of total EHEC ranging from approximately 1 to greater than

17,022 CFUs/100 cm² (95% CI 0.4–72,000 CFUs/100 cm²). For the MPN analysis of total EHEC on the carcasses, 91 samples were included. As observed with the qPCR assay, the MPN assay indicated that the carcass samples have a low concentration of total EHEC, which ranged from approximately 0.1–1 CFU/100 cm² (95% CI 0.02–4 CFUs/100 cm²) from 26 (29%) of the samples.

Comparison of the qPCR and MPN for the Enumeration of Total EHEC from Veal Hide and Carcass Samples

By comparing the qPCR assay to the MPN assay, the qPCR assay was able to enumerate total EHEC in 23 (50%) of the hide samples that were also enumerable with the MPN assay. In addition, the qPCR assay was able to enumerate total EHEC in 7 (7%) of the hide samples that was not enumerable by the MPN assay. Samples with an enumerable amount of total EHEC were within the same log₁₀ value for 10 (43%) samples while the remaining 13 (57%) samples were within approximately one to two orders of magnitude difference between the qPCR and MPN assays (Table 2). Samples only enumerable by either qPCR or the MPN assay were below 3 log₁₀ CFUs/100 cm². Regression analysis between the qPCR and MPN assays for the hide samples was significant (*p* < 0.00001) with a Pearson correlation coefficient of 0.63 (Figure 2A) and the Bland-Altman plot indicates that 92 (97%) of the hide samples were within the 95% confidence interval. This suggests that the two methods are interchangeable for the enumeration of total EHEC from hide samples (Figure 3A). Amongst the carcass samples, the qPCR assay was able to enumerate total EHEC in 10 (11%) samples not enumerable by the MPN assay, while the MPN assay was able to enumerate total EHEC in 21 (23%) samples not enumerable by the qPCR assay. In addition, 5 (5%) samples had a concentration of total EHEC that was enumerable by both assays. The Pearson correlation coefficient for the qPCR and MPN assays on the carcass samples was 0.04, which indicates that a relationship between the methods does not exist (Figure 2B). The Bland-Altman plot supports that the two methods are not interchangeable since less than 95% of the samples were within the confidence interval (Figure 3B). The differences in the ability to enumerate total EHEC between the qPCR and MPN assays could be explained by the methodology and limitations of each assay.

dPCR Analysis of Select Veal Hide and Carcass Samples for the Enumeration of Total EHEC

Additional analysis of the total EHEC enumeration observations was performed on select hide and carcass samples using dPCR. To determine the capabilities of dPCR, a separate eight log standard curve from approximately 8.19–1.19 log₁₀ CFUs/mL was created using the EDL 932 reference strain. From this standard curve, it was found, for both targets, that the dPCR assay was within the same log value as the expected inoculums for dilutions containing approximately 3–7 log₁₀ CFUs/mL (Table 3). At the dilution with an expected 8.19 log₁₀ CFUs/mL, the dPCR assay indicated approximately 7.26 log₁₀ CFUs/mL based on the *eae* target while the *ecf1* target indicated 7.44

TABLE 1 | Average C_q values and coefficients of variability from pooled standard curves collected during the enumeration of total EHEC from veal hides and carcasses.

log ₁₀ CFUs/mL	<i>eae</i>		<i>ecf1</i>	
	C _q ± SD ^a	CV ^b	C _q ± SD	CV
7.69	21.09 ± 0.09	5.66%	18.27 ± 0.04	2.59%
6.69	24.37 ± 0.11	6.99%	21.55 ± 0.10	6.49%
5.69	27.54 ± 0.04	2.78%	24.64 ± 0.05	3.24%
4.69	30.83 ± 0.12	7.28%	27.87 ± 0.11	6.95%
3.69	34.18 ± 0.41	25.21%	31.24 ± 0.50	30.85%

^aThe average C_q value ± the standard deviation.

^bCV, coefficient of variability

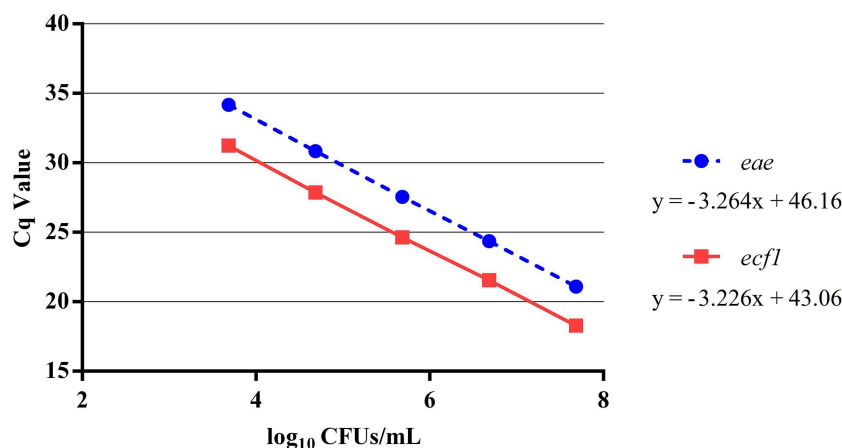


FIGURE 1 | Five log standard curve for *eae* and *ecf1* that was generated from the average of the pooled Cq values. Individual aliquots of the same five log standard curve of the EDL 932 gDNA were loaded in duplicate for each plate during the enumeration of the total EHEC concentration in the paired veal hide and carcass samples. The Cq values for *eae* and *ecf1* at each dilution were pooled and averaged and plotted against the representative EDL 932 concentration. Error bars indicate the average standard deviation.

TABLE 2 | Distribution of total EHEC enumerated from paired veal hides and carcasses using qPCR and MPN assays.

log ₁₀ CFUs/100cm ²	Hides		log ₁₀ CFUs/100cm ²	Carcasses	
	qPCR (%)	MPN (%)		qPCR (%)	MPN ^a (%)
4	1 (1)	4 (4)	2	4 (4)	0 (0)
3	9 (9)	5 (5)	1	11 (12)	0 (0)
2	19 (20)	15 (16)	<1	1 (1)	26 (29)
1	1 (1)	22 (23)	NE ^b	79 (83)	65 (71)
NE ^b	65 (69)	49 (52)			

^an = 91.

^b NE, No enumeration.

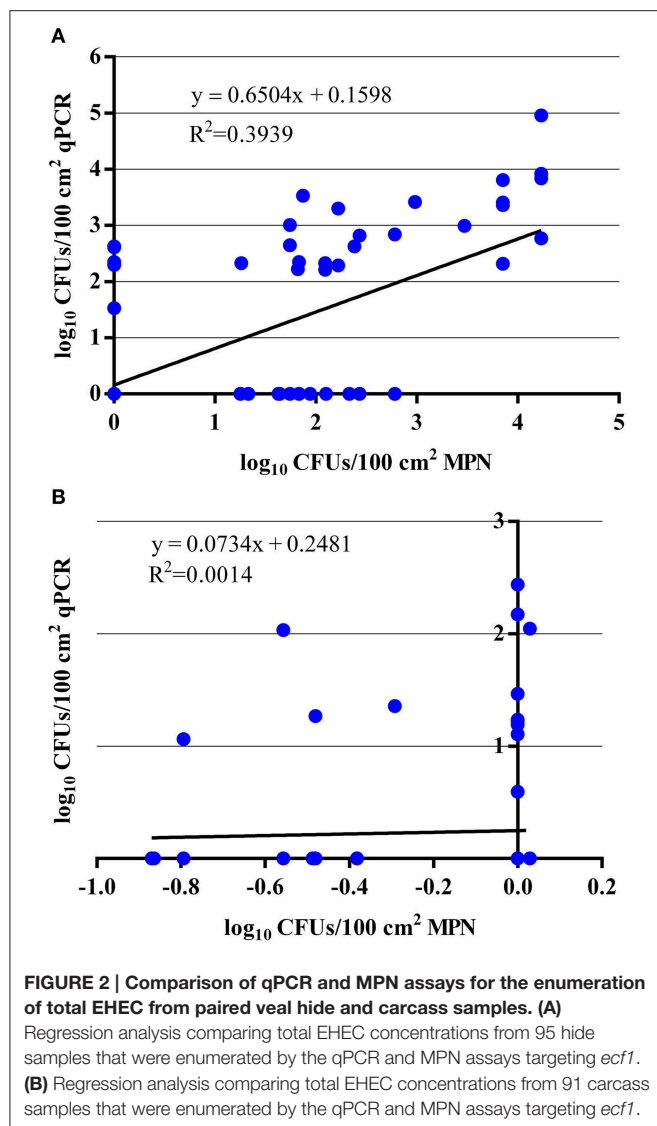
log₁₀ CFUs/mL. Thus, the upper limit of this dPCR assay is approximately within the 7 log₁₀ CFUs/mL range. Moreover, the lower limit was found to be approximately 3 log₁₀ CFUs/mL since at the expected dilution of 2.19 log₁₀ CFUs/mL the concentration of the *eae* and *ecf1* targets was 3.52 log₁₀ and 3.15 log₁₀ CFUs/mL, respectively (Figure 4). Moreover, at the expected dilutions of 2.19 and 1.19 log₁₀ CFUs/mL, the precision for both targets was above 100% (Table 3). Analyzing the same standard curve template DNA by qPCR showed a similar trend, from dilutions containing approximately 7.19 to 3.19 log₁₀ CFUs/mL, as the dPCR. The respective efficiency for *eae* and *ecf1* over the five log curve was 92% and 94%, with an R² of 0.999 for both targets (Supplementary Figure 1).

The dPCR assay tended to overestimate the concentration of total EHEC in the select hide and carcass samples that were indicated previously by the qPCR and MPN assays to have less than 3 log₁₀ CFUs/100 cm² (Figures 5A,B). However, for the hide samples with total EHEC over 3 log₁₀ CFUs/100 cm² the dPCR assay was in the same order of magnitude for 5 (50%) and 6 (60%) of the samples enumerated by either qPCR or the

MPN assay, respectively. In three samples, 4, 49, and 62, the dPCR assay estimated the total EHEC concentration closer to the concentration enumerated by the MPN assay while the qPCR assay determined the EHEC concentration to be greater than approximately one magnitude lower (Supplementary Table 1). In addition, the samples 45, 51, and 53 were closer in the estimated concentration between the qPCR and dPCR assay than the MPN assay (Supplementary Table 1). The carcass samples selected for dPCR analysis all had total EHEC below 3 log₁₀ CFUs/100 cm² as determined by the qPCR and MPN assays. Despite the low level of total EHEC, the dPCR assay determined the concentration within the same log₁₀ for approximately 3 (19%) and was an order of magnitude higher or lower for 10 (62%) and two orders higher for 3 (19%) of the samples as the qPCR assay (Supplementary Table 2). The MPN assay estimated the total EHEC concentration on the carcasses at two to three orders of magnitude lower than the qPCR and dPCR assays (Supplementary Table 2).

Prevalence of Total EHEC in Veal Hide and Carcass Enrichments

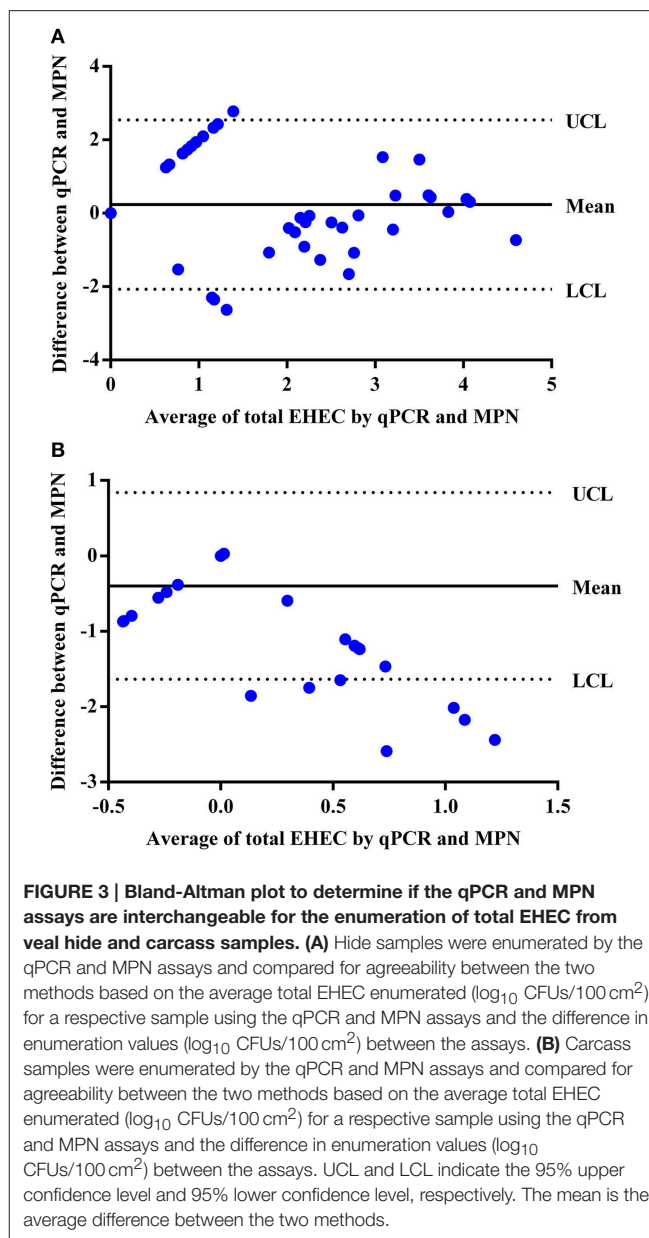
To determine the prevalence of total EHEC in the hide and carcass samples, the samples were enriched and a multiplex qPCR assay targeting *eae*, *ecf1*, and *stx1/2* was performed. From the hide samples, total EHEC, which are positive for all three targets, was detected in 72 (76%) of the samples while 5 (5%) samples were positive for only *eae* and *ecf1*. In addition, *eae* and *stx1/2* was detected in 12 (13%) of the samples while *eae* alone was detected in 6 (6%) samples. Using a χ^2 test, significantly ($p < 0.05$) more carcass samples were positive for all three targets than the hide samples with 86 samples. This accounted for approximately 91% of the carcass samples while 3 (3%) samples had detectable amounts of only *eae* and *ecf1*. Four (4%) of the carcass samples had detectable amounts of only *eae* and *stx1/2* and 1 (1%) sample contained only *eae* while 1 (1%) sample was negative for all three targets. The average post enrichment Cq values for the



hide and carcass samples that were positive for all three targets were compared. The hide samples had an average Cq value \pm the standard deviation for *eae*, *ecf1*, and *stx1/2* of 27.4 ± 1.6 , 30.8 ± 2.7 , and 29.0 ± 3.0 respectively. For the carcass samples, the respective average Cq values were 33.4 ± 2.9 , 34.7 ± 1.9 , and 34.2 ± 3.1 for *eae*, *ecf1*, and *stx1/2*. A correlation between the pre-enrichment enumeration values and the post-enrichment Cq values for hide and carcass samples was not identified (data not shown).

Discussion

Pathogenic *E. coli* remains a constant concern for food safety and human health with an emphasis on the most severe pathotype, EHEC (Palaniappan et al., 2006). A systematic review and meta-analysis covering 62 years of published reports indicates a stable and continued association of EHEC with calves (Kolenda et al., 2015), yet veal calves have received limited study toward EHEC detection and enumeration methods (Wang et al., 2014). To



address these issues, we sought to investigate the use of qPCR, molecular MPN, and dPCR assays to detect and enumerate total EHEC from paired pre-intervention veal calf hides and carcasses.

Studies utilizing qPCR to detect and enumerate EHEC primarily focus on a single EHEC such as O157:H7 or target virulence genes that can be independently possessed by non-EHEC in a polymicrobial matrix. This use of potential non-conjoined targets results in false positives and an over estimation of the true EHEC population (Jacob et al., 2012). Recently, Livezey et al. (2015) reported the use of *ecf1* as a target for the detection of total EHEC in beef samples. That study detailed the specificity of *ecf1* in *E. coli* possessing *eae*, *stx1/2*, and *ehxA*, which is applicable to determining the total EHEC load in a sample. In addition, qPCR has primarily been used to detect and enumerate specific EHEC from cattle feces, while the application

of qPCR on direct hide and carcass samples is unreported. This is likely due to the low concentration of a specific EHEC serogroup within a defined area on the hide or carcass (Arthur et al., 2007), and the intrinsic limit of detection and enumeration for qPCR, which can range between 10^3 and 10^4 CFUs/mL. Indeed, based on the surface area sampled in this study, to reach the theoretical limit of detection and enumeration comparable to the previously determined reliable limit of 1250 CFUs/mL (Luedtke et al., 2014) a concentration of 5000 and 417 CFUs/100 cm² is required for the hide and carcass samples, respectively.

While we could not find any published reports of using qPCR to determine the total EHEC concentration on cattle or veal hides or carcasses, a PCR-MPN based investigation of potential total EHEC from 11 head of cattle found the respective average concentration on hides and carcasses to be approximately 15662 and 123 CFUs/100 cm² when using the average of the targets *eae* and *ehxA* for the enumeration regardless of the cattle diet (Gilbert et al., 2008). Analyzing the data in this manner provides a better normalization to our enumeration targeting *ecf1* as *ecf1* is in a 1:1 relationship with *E. coli* possessing both *eae* and *ehxA* (Livezey

et al., 2015). Using this methodology suggests that our use of qPCR to enumerate total EHEC from veal hides and carcasses may under estimate the total EHEC load (Gilbert et al., 2008). Moreover, determining the precise amount of total EHEC in low concentration samples is difficult due to the Monte Carlo effect (Bustin and Nolan, 2004) as respectively 87 and 100% of the enumerable hide and carcass samples were below the reliable limit of enumeration and some samples returned a single Cq value for the duplicate reactions.

Variations of the MPN assay have been described. These modifications incorporate a combination of immunomagnetic separation and PCR to increase sensitivity. However, these MPN assays primarily focus on the enumeration of O157 and O26 from feces (Widiasih et al., 2004; Stephens et al., 2007; Guy et al., 2014) while two studies have investigated the total potential EHEC in feces and on hides and carcasses (Gilbert et al., 2005, 2008). With our dilution range for the hide and carcass samples the enumeration of total EHEC was mostly one to two logs below the estimates of average potential EHEC reported by Gilbert et al. (2008). However, Gilbert et al. (2008) was analyzing cattle and differences in cattle versus veal EHEC hide carriage and carcass processing techniques may exist in addition to physiological and environmental differences (Cristancho et al., 2008; Wang et al., 2014). In addition, the use of an *a priori* dilution scheme, like used here, for the MPN assays highlights the limitations in the ability to enumerate total EHEC at the upper and lower concentrations from a diversity of samples. To encompass a diverse sample set by expanding the MPN dilution range would reduce the throughput and increase the cost per sample of large analyses.

Comparatively, the MPN assay was able to detect and enumerate total EHEC from 17 (17%) more hide and 11 (12%) more carcass samples than the qPCR assay. However, the qPCR assay was able to enumerate total EHEC in 7 hide and 11 carcass samples that were not enumerable by the MPN assay. The additional samples enumerable by PCR could be due to differences in the detection methods. In the qPCR assay, all amplifiable DNA contributes to the enumeration, which would

TABLE 3 | Expected reference strain input and returned output for concentration of *eae* and *ecf1* targets from dPCR assay.

Input log ₁₀ CFUs/mL	<i>eae</i>		<i>ecf1</i>	
	Output log ₁₀ CFUs/mL	dPCR precision (%)	Output log ₁₀ CFUs/mL	dPCR precision (%)
8.19	7.26	1.84	7.44	2.18
7.19	7.20	1.80	7.39	2.02
6.19	6.18	3.70	6.35	3.11
5.19	5.55	7.50	5.52	7.71
4.19	4.64	22.40	4.50	26.83
3.19	3.95	56.78	3.67	85.86
2.19	3.52	109.76	3.15	210.06
1.19	3.38	140.26	2.99	299.85

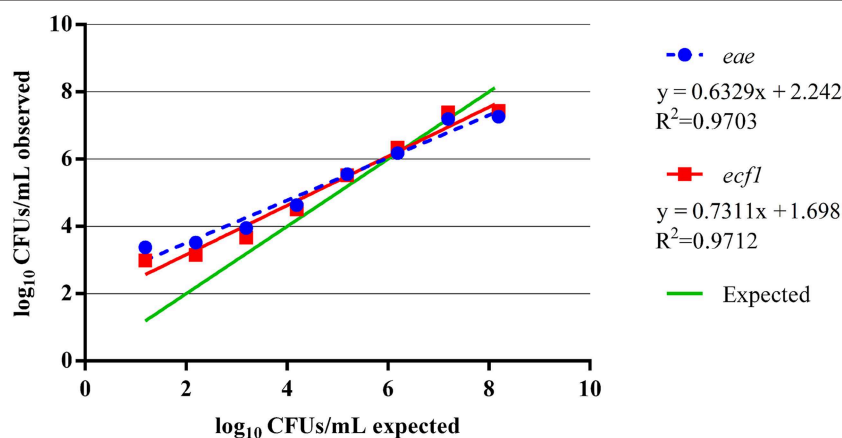


FIGURE 4 | Performance of the dPCR assays over an eight log standard curve. A separate eight log standard curve was created using gDNA from the EDL 932 reference strain. The standard curve was run on a single dPCR chip and targeted *eae* and *ecf1* and plotted against the expected input target concentration.

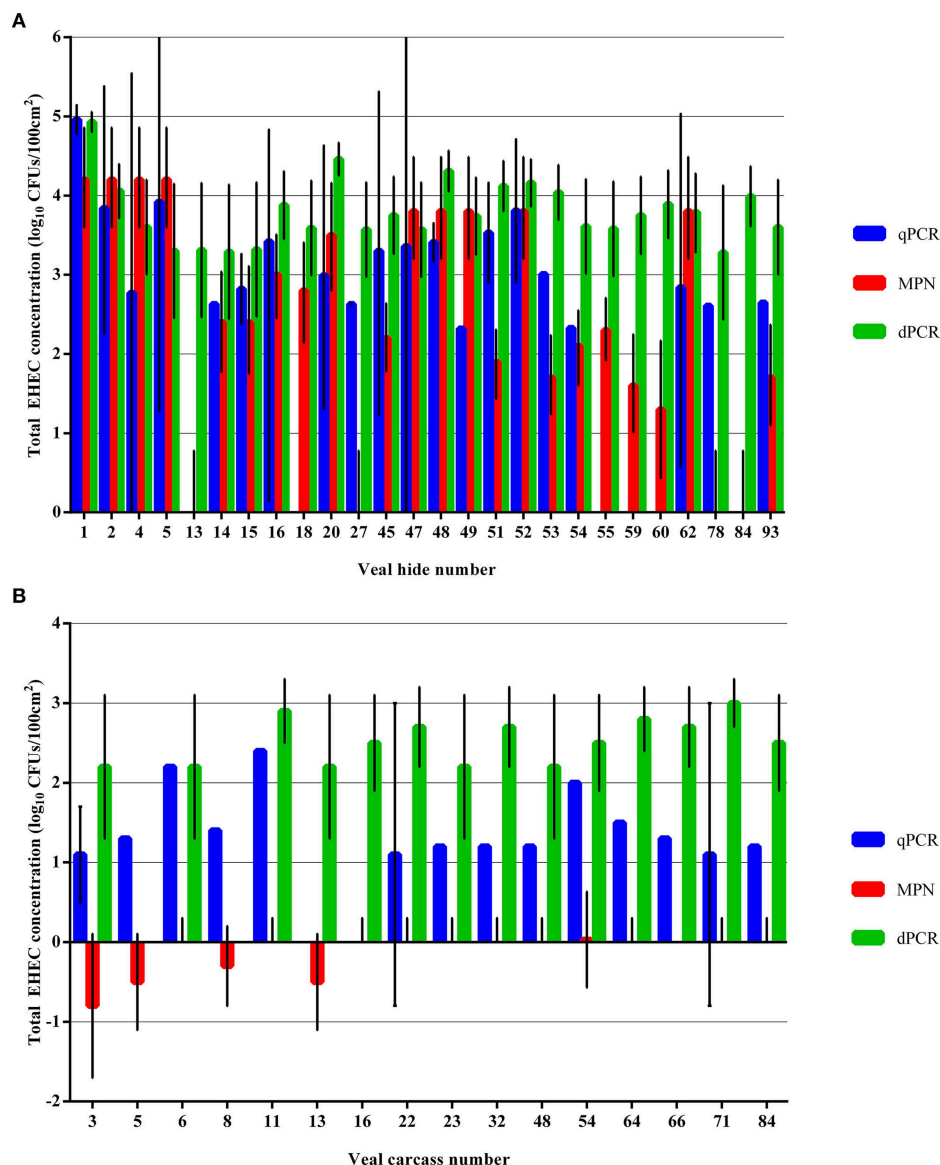


FIGURE 5 | Comparison of total EHEC estimated from select veal hide and carcass samples using qPCR, MPN, and dPCR assays targeting *ecf1*. (A) Total EHEC from selected hide samples were enumerated using the dPCR assay and compared to the enumeration values returned by the qPCR and MPN assays. (B) Total EHEC from selected carcass samples were enumerated using the dPCR assay and compared to the enumeration values returned by the qPCR and MPN assays. Error bars indicate the 95% confidence interval for each sample. A 95% confidence interval was not calculated for samples with a single Cq value from the qPCR assay.

include free DNA and DNA from injured *E. coli*, while the MPN assay targets *ecf1* mRNA transcripts and would likely only enumerate viable cells. In addition, the targeting of mRNA would improve the sensitivity of the assay since more template would be available for detection, which is likely why more hide and carcass samples were enumerable than with the qPCR assay. Despite the detection and enumeration advantage of the MPN assay, the indication, from the Bland Altman plot, that the MPN and qPCR assay are interchangeable for hide enumerations would save a considerable amount of time and resources when using the qPCR assay. The reliable detection and enumeration of total EHEC

from carcasses offers an additional challenge due to the inherently low concentrations of total EHEC that resulted in a difference of 2 to 3 orders of magnitude between the qPCR and MPN assays.

To overcome the low concentrations of total EHEC, we sought to investigate dPCR. Digital PCR has been previously shown to be insensitive to PCR inhibitors and perform similarly to qPCR on environmental samples, but does not require a standard curve to enumerate the target DNA concentration (Blaya et al., 2015; Kinz et al., 2015). Our analysis of a dilution curve using dPCR showed a similarity to qPCR in the enumeration of the gene

targets with the expected input of approximately 10^7 and 10^3 CFUs/mL being within the same order of magnitude for the two methods. However, based on precision, the dynamic range of the dPCR was limited compared to the qPCR assay. To increase the dynamic range toward lower target concentrations, additional dPCR replicates are required to lower the percent precision within an acceptable range (Blaya et al., 2015; Majumdar et al., 2015). The requirement to perform additional dPCR reactions for a single sample with a high percent precision value could be cost inhibitory for commercial application; hence additional dPCR reactions were not performed in this study.

False positive and negative reactions at lower target concentrations can impact the accurate enumeration of the target (Majumdar et al., 2015) as can incidental that pipetting errors between dilutions that change the expected concentration, which the absolute enumeration of dPCR would detect and provide a true estimation (Kishida et al., 2014). dPCR is less prone to error due to stochastic effects like qPCR at low target concentrations. This was evident in hide samples 4, 49, and 62 as the qPCR assay had a high standard deviation between duplicates while the dPCR was in agreement with the concentration estimated by the MPN assay. To our knowledge, this is the first report of using dPCR to enumerate total EHEC from an environmental source and without using previous DNA purification methods.

Our analysis of the 95 hide samples after enrichment indicted total EHEC prevalence at 76%, which is below the 94% average prevalence of the Top 6 EHEC on 132 veal hides as reported by Wang et al. (2013). This difference in prevalence could be due to the origin of the calves and/or the detection method utilized. The method used by Wang et al. (2013) identifies *stx1/2* and single-nucleotide polymorphisms associated with the Top 6 serogroup and *eae*. However, the analysis of results from a polymicrobial sample like hides could result in false positives as we observed 95 (100%) of the hide samples possessing *eae* while 18 (19%) lacked *ecf1* and 12 (67%) of these samples also contained *stx1/2*. This would cause a misinterpretation of true positives if all targets were possessed by separate bacteria, and has been indicated to occur, using the same method, during the detection of the Top 7 on beef hides (Stromberg et al., 2015). Interestingly, the carcass samples had significantly more ($P < 0.05$) samples possessing all three gene targets. With these total EHEC detections coming from an enrichment of the sample, it suggests that the hide intervention utilized maybe not be effective and/or the equipment and procedures utilized for hide removal facilitate further total EHEC transmission (United States Department of Agriculture, and Food Safety and Inspection Service, 2013). However, the comparison of Cq values, although not a fully quantitative method, between the hide and carcass sample sets does provide a generalized estimation of the total EHEC population prior to enrichment with regards to the background microflora population and enrichment media (Vimont et al., 2007). With this methodology, the pre-enrichment hide samples likely started with a total EHEC near the detection/enumeration limit while the carcass samples were below the limit (Luedtke et al., 2014) as we observed for the qPCR, MPN, and dPCR assays. Moreover, differences in the background microflora between samples would explain why a

correlation between the enumeration value and the Cq value was not identified (Vimont et al., 2007). In addition, samples possessing *eae* and *ecf1*, which are classified as atypical EPEC (Livezey et al., 2015), remain a concern due to the potential for these cells to regain *stx1/2* at later point (Bielaszewska et al., 2007) but would not be identified using the conventional detection and enumeration methods.

In conclusion, veal has received limited research pertaining to food safety despite being a significant source of total EHEC. This study is the first to enumerate total EHEC from paired pre-intervention veal hides and carcasses. Each of the methods utilized for detection and enumeration had benefits and drawbacks. The qPCR assay was easy to use and offered a high throughput, but the inherent low concentration of total EHEC on the hides and carcasses limited the accuracy and the requirement of a standard curve limits consistency and number of samples loaded. Our MPN assay allowed for the detection of viable cells and offers a lower limit of detection; however, this sample size inhibited the throughput and required a different range of dilutions. Digital PCR offers advantages of a medium throughput and no standard curve, although the dynamic range of dPCR is limited and to improve precision would require additional analysis of samples with a total EHEC concentration below the reliable limit. When attempting to enumerate low concentrations of total EHEC, the MPN assay would provide the most accurate results while the qPCR and dPCR assays would be effective in determining veal with hide concentrations above approximately 5000 CFUs/100 cm² within 4 h. By rapidly determining highly contaminated calves, these animals could be restricted to end of the day production or receive an increased focus during intervention. Moreover, it was unexpected to detect a higher prevalence of total EHEC on the carcass samples, but the use of an additional carcass intervention may eliminate the risk imposed by the low concentration of total EHEC on the carcass. However, further research tracking the potential spread of total EHEC from a veal hide to the resulting carcasses in combination with our detection and enumeration assays would provide a better understanding of why and how veal trim has an higher association with total EHEC than beef trim.

Author Contributions

BL and JB contributed equally to the designing of experiments, conducting assays, data analysis, and drafting the manuscript.

Acknowledgments

The authors would like to thank the cooperating veal processing plant for access to sample collection and Roka Bioscience for providing the assay tubes and performing analyses. Additional recognition to Lawnie Luedtke and Greg Smith for technical support and Jody Gallagher for secretarial assistance.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01062>

References

- Almanza, A. V. (2011). "Shiga toxin-producing *Escherichia coli* in certain raw beef products," in *Federal Register*, Vol. 77, 9 CFR Parts 416, 417, and 430 Edn. (Washington, DC: United States Department of Agriculture & Food Safety and Inspection Service), 31975–31981.
- Arthur, T. M., Bosilevac, J. M., Brichta-Harhay, D. M., Kalchayanand, N., Shackelford, S. D., Wheeler, T. L., et al. (2007). Effects of a minimal hide wash cabinet on the levels and prevalence of *Escherichia coli* O157:H7 and *Salmonella* on the hides of beef cattle at slaughter. *J. Food Prot.* 70, 1076–1079.
- Arthur, T. M., Keen, J. E., Bosilevac, J. M., Brichta-Harhay, D. M., Kalchayanand, N., Shackelford, S. D., et al. (2009). Longitudinal study of *Escherichia coli* O157:H7 in a beef cattle feedlot and role of high-level shedders in hide contamination. *Appl. Environ. Microbiol.* 75, 6515–6523. doi: 10.1128/AEM.00081-09
- Becker, M. M., and Groschel, B. (2014). *Compositions, Kits, and Related Methods for Detecting and/or Monitoring Shiga Toxin Producing Escherichia coli*. United States Patent US20140342362 A1. Available online at: <https://www.google.com/patents/US20140342362>
- Bielaszewska, M., Prager, R., Köck, R., Mellmann, A., Zhang, W., Tschape, H., et al. (2007). Shiga toxin gene loss and transfer *in vitro* and *in vivo* during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Appl. Environ. Microbiol.* 73, 3144–3150. doi: 10.1128/AEM.02937-06
- Blaya, J., Lloret, E., Santísima-Trinidad, A. B., Ros, M., and Pascual, J. A. (2015). Molecular methods (digital PCR and real-time PCR) for the quantification of low copy DNA of *Phytophthora nicotianae* in environmental samples. *Pest Manag. Sci.* doi: 10.1002/ps.4048. [Epub ahead of print].
- Boerlin, P., McEwen, S. A., Boerlin-Petzold, F., Wilson, J. B., Johnson, R. P., and Gyles, C. L. (1999). Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *J. Clin. Microbiol.* 37, 497–503.
- Bohaychuk, V. M., Gensler, G. E., and Barrios, P. R. (2011). Microbiological baseline study of beef and pork carcasses from provincially inspected abattoirs in Alberta, Canada. *Can. Vet. J.* 52, 1095–1100.
- Brooks, J. T., Sowers, E. G., Wells, J. G., Greene, K. D., Griffin, P. M., Hoekstra, R. M., et al. (2005). Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. *J. Infect. Dis.* 192, 1422–1429. doi: 10.1086/466536
- Bustin, S. A., and Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* 15, 155–166.
- Cristancho, L., Johnson, R. P., McEwen, S. A., and Gyles, C. L. (2008). *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* in white veal calves. *Vet. Microbiol.* 126, 200–209. doi: 10.1016/j.vetmic.2007.06.012
- Elder, R. O., Keen, J. E., Siragusa, G. R., Barkocy-Gallagher, G. A., Koohmaraie, M., and Laegreid, W. W. (2000). Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2999–3003. doi: 10.1073/pnas.97.7.2999
- Gilbert, R. A., Denman, S. E., Padmanabha, J., Fegan, N., Al Ajmi, D., and McSweeney, C. S. (2008). Effect of diet on the concentration of complex Shiga toxin-producing *Escherichia coli* and EHEC virulence genes in bovine faeces, hide and carcass. *Int. J. Food Microbiol.* 121, 208–216. doi: 10.1016/j.ijfoodmicro.2007.11.019
- Gilbert, R. A., Tomkins, N., Padmanabha, J., Gough, J. M., Krause, D. O., and McSweeney, C. S. (2005). Effect of finishing diets on *Escherichia coli* populations and prevalence of enterohaemorrhagic *E. coli* virulence genes in cattle faeces. *J. Appl. Microbiol.* 99, 885–894. doi: 10.1111/j.1365-2672.2005.02670.x
- Goldwater, P. N., and Bettelheim, K. A. (2012). Treatment of enterohemorrhagic *Escherichia coli* (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med.* 10:12. doi: 10.1186/1741-7015-10-12
- Gould, H. (2009). "Update on the epidemiology of STEC in the United States. Non-O157 STEC: waiting for the other shoe to drop. Centers for Disease Control and Prevention," in *Annual Meeting of the Capital Area Food Protection Association* (Washington, DC).
- Guy, R. A., Tremblay, D., Beausoleil, L., Harel, J., and Champagne, M. J. (2014). Quantification of *E. coli* O157 and STEC in feces of farm animals using direct multiplex real time PCR (qPCR) and a modified most probable number assay comprised of immunomagnetic bead separation and qPCR detection. *J. Microbiol. Methods* 99, 44–53. doi: 10.1016/j.mimet.2014.02.002
- Huggett, J. F., Foy, C. A., Benes, V., Emslie, K., Garson, J. A., Haynes, R., et al. (2013). The digital MIQE guidelines: minimum information for publication of quantitative digital PCR experiments. *Clin. Chem.* 59, 892–902. doi: 10.1373/clinchem.2013.206375
- Jacob, M. E., Shi, X., An, B., Nagaraja, T. G., and Bai, J. (2012). Evaluation of a multiplex real-time polymerase chain reaction for the quantification of *Escherichia coli* O157 in cattle feces. *Foodborne Pathog. Dis.* 9, 79–85. doi: 10.1089/fpd.2011.0947
- Jarvis, B., Wilrich, C., and Wilrich, P. T. (2010). Reconsideration of the derivation of Most Probable Numbers, their standard deviations, confidence bounds and rarity values. *J. Appl. Microbiol.* 109, 1660–1667. doi: 10.1111/j.1365-2672.2010.04792.x
- Kinz, E., Leiberer, A., Lang, A. H., Drexel, H., and Muendlein, A. (2015). Accurate quantitation of JAK2 V617F allele burden by array-based digital PCR. *Int. J. Lab. Hematol.* 37, 217–224. doi: 10.1111/ijlh.12269
- Kishida, N., Noda, N., Haramoto, E., Kawaharasaki, M., Akiba, M., and Sekiguchi, Y. (2014). Quantitative detection of human enteric adenoviruses in river water by microfluidic digital polymerase chain reaction. *Water Sci. Technol.* 70, 555–560. doi: 10.2166/wst.2014.262
- Kolenda, R., Burdukiewicz, M., and Schierack, P. (2015). A systematic review and meta-analysis of the epidemiology of pathogenic *Escherichia coli* of calves and the role of calves as reservoirs for human pathogenic *E. coli*. *Front. Cell Infect. Microbiol.* 5:23. doi: 10.3389/fcimb.2015.00023
- Livezey, K. W., Groschel, B., and Becker, M. M. (2015). Use of the *ecf1* gene to detect Shiga toxin-producing *Escherichia coli* in beef samples. *J. Food Prot.* 78, 675–684. doi: 10.4315/0362-028X.JFP-14-417
- Luedtke, B. E., Bono, J. L., and Bosilevac, J. M. (2014). Evaluation of real time PCR assays for the detection and enumeration of enterohemorrhagic *Escherichia coli* directly from cattle feces. *J. Microbiol. Methods* 105C, 72–79. doi: 10.1016/j.mimet.2014.07.015
- Majumdar, N., Wessel, T., and Marks, J. (2015). Digital PCR modeling for maximal sensitivity, dynamic range and measurement precision. *PLoS ONE* 10:e0118833. doi: 10.1371/journal.pone.0118833
- Marx, V. (2014). PCR: paths to sensitivity. *Nat. Meth.* 11, 241–245. doi: 10.1038/nmeth.2849
- Monaghan, A., Byrne, B., Fanning, S., Sweeney, T., McDowell, D., and Bolton, D. J. (2012). Serotypes and virulotypes of non-O157 shiga-toxin producing *Escherichia coli* (STEC) on bovine hides and carcasses. *Food Microbiol.* 32, 223–229. doi: 10.1016/j.fm.2012.06.002
- Palaniappan, R. U., Zhang, Y., Chiu, D., Torres, A., Debroy, C., Whittam, T. S., et al. (2006). Differentiation of *Escherichia coli* pathotypes by oligonucleotide spotted array. *J. Clin. Microbiol.* 44, 1495–1501. doi: 10.1128/JCM.44.4.1495-1501.2006
- Pearce, M. C., Jenkins, C., Vali, L., Smith, A. W., Knight, H. I., Cheasty, et al. (2004). Temporal shedding patterns and virulence factors of *Escherichia coli* serogroups O26, O103, O111, O145, and O157 in a cohort of beef calves and their dams. *Appl. Environ. Microbiol.* 70, 1708–1716. doi: 10.1128/AEM.70.3.1708-1716.2004
- Ramamurthy, T., Ghosh, A., Pazhani, G. P., and Shinoda, S. (2014). Current perspectives on viable but non-culturable (VBNC) pathogenic bacteria. *Front. Public Health* 2:103. doi: 10.3389/fpubh.2014.00103
- Russo, P., Botticella, G., Capozzi, V., Massa, S., Spano, G., and Beneduce, L. (2014). A fast, reliable, and sensitive method for detection and quantification of *Listeria monocytogenes* and *Escherichia coli* O157:H7 in ready-to-eat fresh-cut products by MPN-qPCR. *Biomed Res. Int.* 2014:608296. doi: 10.1155/2014/608296
- Sivaganesan, M., Haugland, R. A., Chern, E. C., and Shanks, O. C. (2010). Improved strategies and optimization of calibration models for real-time PCR absolute quantification. *Water Res.* 44, 4726–4735. doi: 10.1016/j.watres.2010.07.066
- Stephens, T. P., Loneragan, G. H., Chaney, W. E., Branham, L. A., and Brashers, M. M. (2007). Development and validation of a most-probable-number-immunomagnetic separation methodology of enumerating *Escherichia coli* O157 in cattle feces. *J. Food Prot.* 70, 1072–1075.
- Stromberg, Z. R., Baumann, N. W., Lewis, G. L., Severt, N. J., and Cernicchiaro, N., Renter, et al. (2015). Prevalence of enterohemorrhagic *Escherichia coli* O26, O45, O111, O121, O145, and O157 on hides and preintervention carcass surfaces of feedlot cattle at harvest. *Foodborne Pathog. Dis.* 12, 631–638. doi: 10.1089/fpd.2015.1945

- United States Department of Agriculture and Food Safety and Inspection Service. (2012). *Documenting Sample Source when Collecting Samples for Shiga Toxin-Producing Escherichia coli (STEC) Verification Testing, Notice 72-12*. Washington, DC: Department of Agriculture. Available online at: <http://www.fsis.usda.gov/OPPDE/rdad/FSISNotices/72-12.pdf> (Accessed December 9, 2014).
- United States Department of Agriculture, and Food Safety and Inspection Service. (2013). *Increased Verification by Inspection Program Personnel of Sanitary Dressing at Veal Slaughter Establishments, Notice 20-13*. Washington, DC: Department of Agriculture. Available online at: <http://www.fsis.usda.gov/OPPDE/rdad/FSISNotices/20-13.pdf> (Accessed December 9, 2014).
- Vimont, A., Vernozy-Rozand, C., Montet, M. P., Bavai, C., Fremaux, B., and Delignette-Muller, M. L. (2007). Growth of Shiga-toxin producing *Escherichia coli* (STEC) and bovine feces background microflora in various enrichment protocols. *Vet. Microbiol.* 123, 274–281. doi: 10.1016/j.vetmic.2007.02.003
- Wang, F., Yang, Q., Kase, J. A., Meng, J., Clotilde, L. M., Lin, A., et al. (2013). Current trends in detecting non-O157 Shiga toxin-producing *Escherichia coli* in food. *Foodborne Pathog Dis.* 10, 665–677. doi: 10.1089/fpd.2012.1448
- Wang, R., Koohmaraie, M., Luedtke, B. E., Wheeler, T. L., and Bosilevac, J. M. (2014). Effects of in-plant interventions on reduction of enterohemorrhagic *Escherichia coli* and background indicator microorganisms on veal calf hides. *J. Food Prot.* 77, 745–751. doi: 10.4315/0362-028X.JFP-13-388
- Wasilenko, J. L., Fratamico, P. M., Narang, N., Tillman, G. E., Ladely, S., Simmons, M., et al. (2012). Influence of primer sequences and DNA extraction method on detection of non-O157 Shiga toxin-producing *Escherichia coli* in ground beef by real-time PCR targeting the *eae*, *stx*, and serogroup-specific genes. *J. Food Prot.* 75, 1939–1950. doi: 10.4315/0362-028X.JFP-12-087
- Widiasih, D. A., Ido, N., Omoe, K., Sugii, S., and Shinagawa, K. (2004). Duration and magnitude of faecal shedding of Shiga toxin-producing *Escherichia coli* from naturally infected cattle. *Epidemiol. Infect.* 132, 67–75. doi: 10.1017/S0950268803001468
- Wieler, L. H., Schwanitz, A., Vieler, E., Busse, B., Steinruck, H., Kaper, J. B., et al. (1998). Virulence properties of Shiga toxin-producing *Escherichia coli* (STEC) strains of serogroup O118, a major group of STEC pathogens in calves. *J. Clin. Microbiol.* 36, 1604–1607.

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

Copyright © 2015 Luedtke and Bosilevac. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Detection, Characterization, and Typing of Shiga Toxin-Producing *Escherichia coli*

Brendon D. Parsons¹, Nathan Zelyas², Byron M. Berenger² and Linda Chui^{1*}

¹ Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB, Canada, ² Medical Microbiology and Immunology, University of Alberta, Edmonton, AB, Canada

OPEN ACCESS

Edited by:

Pina Fratamico,
Agricultural Research Service, USA

Reviewed by:

Nora Lía Padola,
Universidad Nacional del Centro de la
Provincia de Buenos Aires, Argentina
Glenn Edward Tillman,
United States Department of
Agriculture Food Safety and
Inspection Service, USA

*Correspondence:

Linda Chui
linda.chui@albertahealthservices.ca

Specialty section:

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

Received: 27 February 2016

Accepted: 22 March 2016

Published: 12 April 2016

Citation:

Parsons BD, Zelyas N, Berenger BM
and Chui L (2016) Detection,
Characterization, and Typing of Shiga
Toxin-Producing *Escherichia coli*.
Front. Microbiol. 7:478.
doi: 10.3389/fmicb.2016.00478

Shiga toxin-producing *Escherichia coli* (STEC) are responsible for gastrointestinal diseases reported in numerous outbreaks around the world. Given the public health importance of STEC, effective detection, characterization and typing is critical to any medical laboratory system. While non-O157 serotypes account for the majority of STEC infections, frontline microbiology laboratories may only screen for STEC using O157-specific agar-based methods. As a result, non-O157 STEC infections are significantly under-reported. This review discusses recent advances on the detection, characterization and typing of STEC with emphasis on work performed at the Alberta Provincial Laboratory for Public Health (ProvLab). Candidates for the detection of all STEC serotypes include chromogenic agars, enzyme immunoassays (EIA) and quantitative real time polymerase chain reaction (qPCR). Culture methods allow further characterization of isolates, whereas qPCR provides the greatest sensitivity and specificity, followed by EIA. The virulence gene profiles using PCR arrays and stx gene subtypes can subsequently be determined. Different non-O157 serotypes exhibit markedly different virulence gene profiles and a greater prevalence of stx1 than stx2 subtypes compared to O157:H7 isolates. Finally, recent innovations in whole genome sequencing (WGS) have allowed it to emerge as a candidate for the characterization and typing of STEC in diagnostic surveillance isolates. Methods of whole genome analysis such as single nucleotide polymorphisms and k-mer analysis are concordant with epidemiological data and standard typing methods, such as pulsed-field gel electrophoresis and multiple-locus variable number tandem repeat analysis while offering additional strain differentiation. Together these findings highlight improved strategies for STEC detection using currently available systems and the development of novel approaches for future surveillance.

Keywords: STEC, detection, characterization, typing, O157, non-O157

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) encompass a heterogeneous group of enteric pathogens responsible for numerous sporadic infections and large outbreaks worldwide. Accurate and rapid diagnosis of STEC infections is important for the appropriate management of infected patients and for implementation of proper public health interventions. Specifically, patients infected with STEC should not be treated with antibiotics because of the risk of developing

hemolytic uremic syndrome (HUS) (Wong et al., 2000, 2012; Smith et al., 2012). Also, once STEC is identified in a patient, the contacts and potential sources of infection must be identified to prevent further spread of the disease. Although, laboratories have become proficient at detecting O157:H7 infections, they often do not screen stools for other STEC serotypes. This creates a gap in diagnostics; since 50% or more of STEC infections may be caused by non-O157 STEC, our surveillance and understanding of the epidemiology of STEC disease is incomplete (Fey et al., 2000; Jelacic et al., 2003; Thompson et al., 2005; Chui et al., 2011; Couturier et al., 2011; Scallan et al., 2011; Gould et al., 2013; STEC, National Surveillance Summary, 2014).

Growing recognition of the shortfall in STEC detection has prompted a shift toward more comprehensive STEC identification methods. Bacteriological culture remains the gold standard test, given the importance of identifying viable bacterial isolates for typing. For this reason, there has been increased development and use of agars which also select for non-O157 STEC (Kase et al., 2015). However, as culture-based methods are laborious and exhibit clear limits in sensitivity for STEC detection, it is recommended that laboratories supplement culture-based approaches with other assay types (Gould et al., 2009).

Alongside culture-based STEC testing many laboratories assay for the presence of Shiga toxins (Stx) or the *stx* genes. Shiga toxin was originally referred to as verotoxin for its cytotoxic effect on Vero cells (Konowalchuk et al., 1977). Once Stx was linked to hemorrhagic colitis and HUS, researchers developed cytotoxicity assays to detect Stx from both fecal specimens as well as from enriched stool cultures containing polymyxin B (Karmali et al., 1985). While such laborious cytotoxicity assays remain a method of diagnosis for some laboratories, detection of Stx or the presence of the *stx* genes is now primarily done in clinical laboratories by enzyme immune assay (EIA) of some form, or by polymerase chain reaction (PCR)-based approaches, respectively. These methods can also determine if Stx1 or 2 are present, which adds prognostic value, since there is a well-documented correlation of Stx2 with the clinical severity of STEC infection and the risk of HUS (Schimmer et al., 2008; Soon et al., 2013; Chui et al., 2015a).

While advancements in the sensitivity and speed of STEC detection have direct implication on the diagnosis and treatment of diarrheal illnesses, characterization of STEC isolates beyond serotype or individual virulence factors is required for prevention, control and prediction of STEC infections on a public health scale. The requirement for high resolution typing of STEC is also increasingly necessary given the observed emergence of diverse types of virulent strains (Soon et al., 2013). Current STEC fingerprinting techniques such as pulsed-field gel electrophoresis (PFGE) or multi-locus variable number tandem repeat analysis (MLVA) using the PulseNet International protocol allow comparison of strains from different countries and aid in the epidemiological tracking of STEC infections around the world, especially during outbreak settings (Sabat et al., 2013). While methods such as PFGE and MLVA play a crucial role in current outbreak investigations, the increasingly tractable use of whole genome sequencing (WGS) has garnered

significant interest as a powerful method for typing bacterial pathogens (Chattaway et al., 2016). WGS technologies promise typing resolution at orders of magnitude greater than existing methods. Yet, as technical capabilities improve for STEC typing, new challenges arise surrounding implementation, standardization and management of typing data (Köser et al., 2012; Franz et al., 2014).

Here we present an overview of recent advances and the experiences at the Provincial Laboratory of Public Health (ProvLab) in Alberta Canada with various bacteriological, molecular and genomic strategies for detection and typing of STEC. We assess the benefits and shortcomings of various methods used in the detection and differentiation of STEC. Through evaluation of available systems and opportunities for novel approaches, this review aims to identify improved strategies for STEC identification and surveillance.

DIFFERENTIAL AND SELECTIVE MEDIA

When *E. coli* O157:H7 was first identified as an etiologic agent of hemorrhagic colitis, it was discovered to be unlike most other strains of *E. coli*, because it could not ferment sorbitol (Wells et al., 1983; Pai et al., 1984). This biochemical peculiarity led to the use of sorbitol-MacConkey (SMAC) agar to identify non-sorbitol fermenting *E. coli* in stool of patients with bloody diarrhea (Remis, 1984). This agar differed from typical MacConkey agar by substituting lactose with sorbitol; non-sorbitol fermenting organisms produced white colonies on the medium (Remis, 1984). Early investigations found that SMAC agar displayed acceptable sensitivity, specificity, and negative predictive value (NPV) for *E. coli* O157:H7 detection, but a positive predictive value (PPV) of only 28% (Table 1; March and Ratnam, 1986). Besides its low PPV, other limitations of SMAC include its inability to detect non-O157 STEC as well as sorbitol-fermenting O157 STEC isolates, which can carry the toxigenic *stx* genes and cause outbreaks (Gunzer et al., 1992; Ammon et al., 1999). To improve the detection of STEC, a new chromogenic medium, CHROMagar™ O157, was developed by CHROMagar Microbiology (CHROMagar™ O157 CHROMagar Microbiology, Paris, France, 2013). Through the incorporation of proprietary chromogenic substrates in CHROMagar™ O157 agar, O157 STEC appear mauve while other *E. coli* are blue. Unfortunately, as with SMAC agar, CHROMagar™ O157 is not able to detect most non-O157 STEC (Bettelheim, 1998). The only study directly comparing SMAC agar to CHROMagar™ O157 showed that CHROMagar™ O157 had a lower false-positive rate for colony picks (20%) compared to that of SMAC agar (65%) as well as estimated annual cost-savings of approximately \$76,000 in a large Canadian clinical laboratory (Church et al., 2007). In the same study, CHROMagar™ O157 had a higher sensitivity (96.3% vs. 85.2%) for detecting O157 STEC (Table 1; Church et al., 2007). Other chromogenic media, such as Colorex™ O157 (Alere, Inc., Ottawa, ON, Canada), are also available to detect O157 STEC, but have not been clinically evaluated.

Many clinical microbiology laboratories continue to use either SMAC agar or CHROMagar™ O157 exclusively to detect

TABLE 1 | Summary of clinically evaluated STEC detection methods.

Assay	Description	Sensitivity/Specificity	PPV/NPV	References
SOLID MEDIA				
CHROMagar™ O104 STEC	Detects O104:H4 STEC through chromogenic substrate	71.4%/99.1%	62.5%/99.4%	Gouali et al., 2013
CHROMagar™ O157	Detects O157 STEC through a chromogenic substrate	96.3%/100%	100%/100%	Church et al., 2007
CHROMagar™ STEC	Detects O157 and non-O157 STEC through a chromogenic substrate	84.6–85.7%/87–95.8%	13.9–60%/98.9–99.6%	Wylie et al., 2013; Zelyas et al., 2016
Rainbow® Agar O157	Detects O157 and some non-O157 STEC through β -glucuronidase and β -galactosidase activities	Not determined	Not determined	Zelyas et al., 2016
Sorbitol-MacConkey	Detects O157 STEC by lack of sorbitol fermentation	85.2–100%/85–100%	28–100%/99.9–100%	March and Ratnam, 1986; Church et al., 2007
TOXIN DETECTION				
BioStar SHIGATOX	Optical immunoassay detects and differentiates Stx1/2	96.8%/99.4%	88.2%/99.9%	Church et al., 2007
Duopath Verotoxin-test™	Immunochromatographic assay detects and differentiates Stx1/2	88.9%/71.4%	42.1%/96.5%	Grif et al., 2007
Immunocard STAT!®	Immunochromatographic assay detects and differentiates Stx1/2	35.5%/99.4%	54.5%/98.6%	Chui et al., 2013
Premier® EHEC	Microwell EIA detects Stx1/2 without toxin differentiation	90.5–96.8%/98.5–100%	76.2–100%/98–99.9%	Grif et al., 2007; Teel et al., 2007; Chui et al., 2011; Hermos et al., 2011
ProSpect™ Shiga Toxin <i>E. coli</i>	Microwell EIA detects Stx1/2 without toxin differentiation	76.8%/100%	100%/96.1%	Gerritzen et al., 2011
Shiga Toxin Chek™	Microwell EIA detects Stx1/2 without toxin differentiation	80%/98.2%	53.3%/99.5%	Chui et al., 2015b
Shiga Toxin Quik Chek™	Immunochromatographic assay detects and differentiates Stx1/2	85%/100%	100%/99.6%	Chui et al., 2015b
MOLECULAR DETECTION				
EntericBio real-time Gastro Panel I®	Multiplex PCR detects four pathogens including STEC	100%/99.8%	82.9%/100%	Koziel et al., 2013
FilmArray® GI panel	Array-based multiplex PCR detects 22 pathogens including STEC	100%/99.1–99.6%	86.8–95.1%/100%	Khare et al., 2014; Buss et al., 2015
Seeplex® Diarrhea ACE Detection	Multiplex PCR with dual priming oligonucleotide technology detecting 14 pathogens including STEC	100%/99.6–100%	66.7–100%/100%	Coupland et al., 2013; Onori et al., 2014
TaqMan® in-house STEC detection assay	Real-time PCR assay developed by Alberta ProVLab detecting <i>stx</i> ₁ and <i>stx</i> ₂ separately	100%/100%	100%/100%	Chui et al., 2010
xTag® Gastrointestinal Pathogen Panel	Multiplex PCR with bead hybridization detecting 15 pathogens including STEC	93.5–100%/98.8–100%	75–100%/99.1–100%	Claas et al., 2013; Navidad et al., 2013; Deng et al., 2015

PPV, positive predictive value; NPV, negative predictive value; for CHROMagar™ O104 STEC, diagnostic values are for O104:H4 STEC only; for CHROMagar™ O157 and sorbitol-MacConkey, diagnostic values are for O157:H7 STEC only; for all other assays, diagnostic values are for all STEC serotypes.

STEC even though they are not appropriate for the detection of non-O157 STEC. Guidelines released by the Centers for Disease Control and Prevention (CDC) specify that laboratories are to simultaneously culture stool specimens for O157 STEC and test them with an assay that detects non-O157 STEC (Gould et al., 2009). Because of the shortcomings of SMAC and CHROMagar™ O157, there has been interest in creating a medium capable of detecting non-O157 serotypes of STEC.

During the 2011 *E. coli* O104:H4 outbreak in Europe, a medium specifically designed to detect the outbreak strain, CHROMagar™ STEC O104, was developed (Gouali et al., 2013). While this agar is able to detect the O104:H4 strain expressing an extended-spectrum β -lactamase (ESBL) initially causing the

outbreak, it is unable to detect other non-O157 serotypes and O104:H4 isolates that have lost the plasmid encoding the ESBL (Mariani-Kurkdjian et al., 2011; Grad et al., 2012). Its utility is also hampered by low sensitivity and PPV (Table 1).

Other chromogenic agars capable of detecting wider ranges of STEC serotypes have been described, including Rainbow® Agar O157 and CHROMagar™ STEC (Biolog, Hayward, CA, USA, 2008; CHROMagar™ STEC, CHROMagar Microbiology, Paris, France, 2014). CHROMagar™ STEC is meant to detect all STEC serotypes. Like other media developed by CHROMagar Microbiology, pathogen detection on CHROMagar™ STEC is based on the organism's utilization of proprietary chromogenic substrates. CHROMagar™ STEC is able to detect most of the

STEC serotypes for which it has been assessed (Hirvonen et al., 2012; Wylie et al., 2013; Zelyas et al., 2016). Direct inoculation of stool onto the agar yields acceptable sensitivity, specificity, and NPV, but the PPV is quite low (Wylie et al., 2013; Zelyas et al., 2016). Two studies using different broth enrichment protocols prior to inoculation showed sensitivities varying from 50% (McCallum et al., 2013) to 91.4% (Gouali et al., 2013).

Rainbow[®] Agar O157 is purported to detect O157:H7, O26:H11, O48:H21, O111:H-, and O111:H8 serotypes based on their reduced or absent β -glucuronidase activity compared to non-toxicogenic strains (Biolog, Hayward, CA, USA, 2008). Although, Rainbow[®] Agar O157 has been evaluated in a number of studies for the detection of STEC in food and water (Radu et al., 2000; Tutenel, 2003; Tillman et al., 2012; Yoshitomi et al., 2012; Ngwa et al., 2013), its ability to identify STEC from human stool was first investigated at the Alberta ProvLab. A study by Zelyas et al. (2016) performed at the Alberta ProvLab compared four chromogenic agar media in their ability to detect non-O157 STEC. Isolates from a panel of 161 non-O157 STEC were inoculated directly onto CHROMagar[™] STEC, Rainbow[®] Agar O157, CHROMagar[™] O157, and Colorex[®] O157 to observe if the isolates would produce STEC-like colonies. Unsurprisingly, CHROMagar[™] O157 and Colorex[®] O157 were unable to identify the majority of non-O157 isolates as STEC, while CHROMagar[™] STEC and Rainbow[®] Agar O157 had detection rates of 90% and 70%, respectively. Using stool cultures spiked with non-O157 STEC isolates, it was found that CHROMagar[™] STEC once again exhibited a superior detection rate of 72% (compared to 26% using Rainbow[®] Agar O157) in bloody stool. Similar to previous studies, CHROMagar[™] STEC demonstrated a sensitivity, specificity, PPV, and NPV of 84.6, 87, 13.9, and 99.6%, respectively, when 536 clinical specimens were inoculated directly onto the medium (Table 1). Although, studies demonstrate that CHROMagar[™] STEC shows promise in its ability to rule out STEC in its absence, the high number of false-positive results seen on the medium would necessitate considerable additional laboratory testing to confirm or deny STEC status of mauve colonies. The use of CHROMagar[™] STEC should perhaps be limited to the procurement of STEC isolates when a stool tests positive for STEC by a non-culture method, such as toxin or toxin gene detection. As discussed below, such non-culture methods often display sensitivities above the ~85% seen with CHROMagar[™] STEC.

ENZYME IMMUNOASSAYS

As no culture medium is yet available for the practical detection of all STEC serotypes, identifying the Shiga toxin (Stx) in stools is an alternative method of diagnosing STEC-related disease. The first EIAs developed for Stx identified STEC colonies based on the binding of monoclonal antibodies to Stx1 and Stx2 immobilized on membranes (Perera et al., 1988; Milley and Sekla, 1993). Since the creation of these early EIAs that required the growth of isolated colonies on a solid medium, a number of other assays have been developed for the detection of Shiga toxin directly from stool or from enriched stool cultures.

One of the most evaluated and used EIAs is the Premier[®] EHEC microwell immunoassay (Meridian Bioscience Inc.,

Cincinnati, OH, USA). Multiple studies using overnight broth enrichment stool cultures found that Premier[®] EHEC demonstrates high sensitivity and specificity (Table 1). Premier[®] EHEC has also been used to detect Stx directly from clinical specimens without the use of an overnight enrichment step; one group found this approach had a sensitivity of 83.9% and specificity of 99.8% (Teel et al., 2007). Another microwell immunoassay that has undergone clinical evaluation, the ProSpect[™] Shiga Toxin *E. coli* assay (Remel, Lenexa, KS, USA), demonstrated inferior sensitivity compared to Premier[®] EHEC (Table 1).

Besides microwell EIAs, other types of immunoassays have been developed to detect STEC. One such assay is the BioStar[®] SHIGATOX optical immunoassay (Inverness Medical Professional Diagnostics, Inc., San Diego, CA, USA) which detects Stx by its interaction with anti-Stx antibodies on the surface of a silicon wafer; this interaction causes an increase in the optical thickness of the thin film and results in a visible color change on the wafer. Similarly, the Duopath Verotoxin-test[™] (Merck, Darmstadt, Germany) is an immunochromatographic assay that employs anti-Stx antibodies immobilized to a membrane to bind and detect Stx. In previous studies, the BioStar[®] SHIGATOX assay exhibited a superior performance to the Duopath Verotoxin-test[™] (Table 1). However, the Duopath Verotoxin-test[™] is advantageous because it differentiates between Stx1- and Stx2-producing STEC.

Studies performed at the Alberta ProvLab have evaluated two microwell immunoassays: the aforementioned Premier[®] EHEC and the Shiga Toxin Chek[™] assay (TechLab, Inc., Blacksburg, VA, USA; Chui et al., 2011, 2015b). Premier[®] EHEC demonstrated a sensitivity of 90.5%, similar to that seen in previous studies (Grif et al., 2007; Teel et al., 2007; Hermos et al., 2011), and the Shiga Toxin Chek[™] assay had a lower sensitivity of 80% which decreased to 70% when unenriched specimens were used (Table 1; Chui et al., 2011, 2015b).

Additionally, two immunochromatographic assays have been assessed at the Alberta ProvLab: ImmunoCard STAT![®] (Meridian Bioscience, Inc., Cincinnati, OH, USA) and Shiga Toxin Quik Chek[™] (TechLab, Inc., Blacksburg, VA, USA; Chui et al., 2015b, 2013). Despite having a specificity >99%, ImmunoCard STAT![®] had a low sensitivity of 35.5% even when using enrichment broths (Table 1). Shiga Toxin Quik Chek[™] demonstrated sensitivities of 85 and 70% with and without enrichment, respectively (Chui et al., 2013, 2015b).

Some caution must be exercised when using EIAs alone to detect STEC. There have been two norovirus outbreaks in the United States in which EIAs yielded false-positive STEC results, highlighting the pitfall of depending on a single method to diagnose STEC-related disease (Centers for Disease Control and Prevention (CDC), 2001, 2006).

MOLECULAR METHODS

While the detection of Stx is a direct way to determine if clinical specimens harbor STEC, there has been much interest in nucleic acid-based methods to detect the presence of the *stx* genes in stools. The earliest application of nucleic acid detection for STEC involved the use of cloned portions of *stx*₁ and *stx*₂ as ³⁵S-labeled

DNA probes in colony hybridization assays (Willshaw et al., 1987; Scotland et al., 1988). Soon after DNA hybridization assays were developed, a conventional PCR targeting *stx*₁ and *stx*₂ in a single reaction was devised (Pollard et al., 1990); a similar assay was later described which could detect STEC from DNA isolated from stool (Brian et al., 1992).

A multitude of PCR assays have been developed since and a number of them use real-time platforms. Some of the advantages of using real-time PCR assays include excellent sensitivity and specificity and the ability to devise multiplex assays to detect and differentiate between *stx*₁ and *stx*₂, other virulence genes such as the intimin gene, *eae*, and hemolysin gene, *ehx4*, and even other gastrointestinal pathogens. The first reported *stx*-targeting real-time PCR assay used directly on naturally infected clinical stool had a sensitivity of 100% and a specificity of 92% (Bélanger et al., 2002). Numerous real-time PCR assays have been designed and generally demonstrate similarly high detection rates with few false positive results (Grys et al., 2009; Gerritzen et al., 2011; Zhang et al., 2012). Commercial real-time PCR assays such as the GeneDisc[®] (GeneDisc[®] Technologies Pall Corporation, NY, USA) and BAX[®] System (DuPont Nutrition and Health, Wilmington, DE, USA) include a panel for rapidly screening for STEC, targeting *stx*₁, *stx*₂, and *eae* or other genes, followed by panels that target serotype-specific genes of O157 STEC and top six non-O157 STEC. These real-time PCR STEC panels exhibit high sensitivity and can be applied in two step screening algorithms that first capture STEC followed by detection the most frequently reported STEC serotypes (Fratamico et al., 2012; Wasilenko et al., 2014). Most real-time PCR assays use any one of a number of available detection systems, including SYBR green, TaqMan[®], molecular beacon probes, fluorescence resonance energy transfer (FRET) probes, LUX[™] (light upon extension) assays with singly-labeled primers without probes, as well as other methods. Further contributing to the heterogeneity of available methods, different real-time PCR assays often target different regions within *stx*₁ and *stx*₂, (Chui et al., 2010).

Numerous multiplex molecular assays for the detection of multiple gastrointestinal pathogens are also available. The xTag[®] Gastrointestinal Pathogen Panel (GPP) (Luminex Corporation, Austin, TX, USA), is FDA- and Health Canada-approved for the detection of multiple agents of gastroenteritis. The GPP employs a multiplex PCR with a reverse transcriptase step intended to amplify nucleic acid from nine bacterial pathogens, three parasites, and three viruses. The generated amplicons are then hybridized to oligonucleotides bound to microspheres, which are detected by the instrument. Included in the GPP are separate targets for the detection of *E. coli* O157 and non-O157 STEC. Multiple evaluations performed in different regions have demonstrated high sensitivities and specificities (Table 1). Some studies report the detection of STEC by the GPP in culture- or conventional PCR-negative specimens; the significance of these results, whether due to heightened sensitivity of the GPP or to false-positives, has not been determined (Mengelle et al., 2013; Vocale et al., 2015). The EntericBio real-time Gastro Panel I[®] (SeroSep, Limerick, Ireland), the FilmArray[®] GI panel (BioFire, Inc., Salt Lake City, UT, USA), and the Seeplex[®] Diarrhea ACE Detection system (Seegene, Seoul, South Korea)

demonstrate similar sensitivities and specificities for STEC as the GPP (Table 1).

A method of considerable interest that has yet to be clinically evaluated for the detection of STEC from human stool is loop-mediated isothermal amplification (LAMP). The assay uses a DNA polymerase with strand-displacement activity and four to six specially-designed primers to generate high numbers of stem-loop amplicons in as little as 1 h at a stable temperature of 60–65°C; real-time visualization of positive reactions occurs with the production of insoluble magnesium pyrophosphate, thus obviating the need for fluorescent reporters (Notomi et al., 2000; Mori and Notomi, 2009). Advantages of LAMP include a high sensitivity and specificity, short turn-around-time, isothermal conditions, and a simple detection method. Two studies employed LAMP to detect STEC from human stool thus far, neither determined sensitivities or specificities for the assays used (Wang et al., 2012; Teh et al., 2014). However, a LAMP assay developed by Hara-Kudo et al. (2007) to detect STEC, had 100% sensitivity for *stx*₁ and *stx*₂, and a specificity of 98% for *stx*₁ and 100% for *stx*₂ in tests of stool samples at the Alberta ProvLab. As well, the PPV for this assay was 92% for *stx*₁ and 100% for *stx*₂ while both *stx*₁ and *stx*₂ had an NPV of 100%. Although, a major disadvantage of LAMP is the difficulty in developing multiplex assays, one could envision how the advantages of the LAMP assay could be exploited in point-of-care testing. However as of yet, more clinical evaluations are needed.

The Alberta ProvLab compared the diagnostic characteristics and costs associated with five PCR assays (Chui et al., 2010). This analysis showed that an in-house assay using the TaqMan[®] platform with a rapid turn-around-time costs the least among the real-time PCR assays, making it the most attractive test (Table 1; Chui et al., 2010). This assay has been used by ProvLab to determine the prevalence of STEC infections in various areas of Alberta during 2006–2012 as well as to act as a comparator for other methods of STEC detection (Chui et al., 2011, 2015b, 2013; Couturier et al., 2011; Zelyas et al., 2016). As well, the Alberta ProvLab is participating in the APPETITE (Alberta Provincial Pediatric Enteric Infection TEam) study, which is comparing the GPP to routine detection methods in a large pediatric cohort from 2014 to 2019 to better define the epidemiology of gastrointestinal disease in Alberta (Freedman et al., 2015). Since routine STEC detection methods in Alberta currently involve only the identification of O157:H7 STEC through culture methods, the use of the GPP during the APPETITE study will greatly enhance STEC disease detection among children with acute gastroenteritis and serve to further evaluate the diagnostic value of the GPP.

AN ALGORITHM TO MAXIMIZE STEC DETECTION

None of the aforementioned approaches is without drawbacks. Culture techniques either lack sensitivity or a robust PPV, toxin detection assays may yield false-positives and are often expensive, and molecular methods tend to be laborious and/or expensive. At the same time, each method has at least one advantage:

culture allows the isolation of strains for typing; EIAs confirm the production of disease-causing toxin and permit non-O157 STEC to be detected; and current nucleic acid tests have high sensitivity and specificity for all STEC serotypes. As suggested by the CDC in 2009, STEC detection algorithms are of most utility if a combination of culture and non-culture methods are used (Gould et al., 2009). One approach would be to pool clinical specimens and test them initially using a non-culture method. This would have the benefit of keeping costs low while screening for a low-prevalence disease. Once a positive result is obtained, the individual stools could be tested by the same non-culture method to identify the STEC-positive specimen (Chou et al., 2014). This would be followed by culture of the specimen on a chromogenic agar to obtain the isolate for typing.

While the design of detection and characterization algorithms may use various combinations of testing methods, the ultimate goal for public health investigations of STEC are aimed at specifically identifying “pathogenic” strains of STEC. However, strategies to detect pathogenic STEC are hindered by a lack of a consistent association between any single marker or combination of markers and the severity of disease. Essentially, there exist no absolute characteristic of pathogenic STEC, and therefore testing algorithms need sufficient inclusivity to capture emerging strains. As outlined by the 2013 European Food Safety Authority (EFSA) criteria for assessing STEC pathogenicity, the O104 outbreak in 2011 revealed significant shortfalls in previous testing algorithms. Specifically, algorithms that focus on identification of a narrow panel of serogroups, virulence genes or reliance on seropathotypes, which define the reported frequencies of certain serotypes with human disease, are likely insufficient to detect “non-typical” emerging STEC strains (EFSA BIOHAZ Panel, 2013). As such, the O104 outbreak strain was not included in any seropathotype category prior to 2011, and screening strategies at the time required detection of *stx* and *eae* before attempting to isolate the suspect STEC, therefore missing O104, which was *eae* negative. Modifications to STEC detection algorithms outlined by the EFSA include requirements to attempt isolation of STEC from all samples positive for *stx* genes. In addition, the EFSA panel also recommended testing for the presence of *aaiC* (a secreted protein of EAEC) and *aggR* (a plasmid-encoded regulator) genes associated with enteroaggregative adhesion, which along with *stx* and *eae* exhibit higher associated risk of severe disease. The continued improvements in STEC identification and strategic testing algorithms will aid epidemiological investigations and provide early detection of future STEC outbreaks.

GENOMICS AND GENOTYPING IN SURVEILLANCE

Once STEC is identified and isolated in culture, the next challenge is to identify the relatedness of isolates for the purpose of public health surveillance. As discussed, STEC isolates can be classified initially based on the serotype, but additional typing is required to determine if an isolate is related to another of the same serotype. PFGE has been used extensively in public health to determine the

relatedness of isolates of many bacterial species including STEC. For STEC, enhanced resolution can be achieved by combining PFGE with MLVA or using PFGE alone. Networks of public health laboratories accredited to run PFGE and/or MLVA report to their regional PulseNet organization the profiles of organisms they type (i.e., PulseNet International). This facilitates identifying national or international outbreaks that would otherwise go unnoticed. In our laboratory, this process through PulseNet Canada and PulseNet USA has found cases linked to Albertan foodborne outbreaks associated with the cross-border trade of food products between different provinces of Canada and the two countries (internal communications).

PFGE and MLVA have great utility in outbreak investigations of STEC and are advantageous because they are amenable to intra-laboratory comparison. However, for many types of bacteria, these methods and others do not have adequate resolution to identify outbreaks. WGS, on the other hand, has been demonstrated in numerous cases to provide enhanced resolution compared to pre-WGS methods (Gilchrist et al., 2015) because the entire genome (or most of the genome) can be analyzed rather than just one or a few genetic elements. In comparison to the current STEC typing and epidemiological screening methods, WGS has superior discriminatory power for comparable cost and would dramatically streamline the detection and typing workflow by replacing the multiple tests required for current investigations (Joensen et al., 2014; Dallman et al., 2015a,b).

WHOLE GENOME SEQUENCING

Before discussing the studies demonstrating the utility of WGS for STEC typing, one must be aware that multiple computing methods exist to assess relatedness between a set of isolates. These methods can be broadly categorized into those that analyze the difference in single nucleotide variants (SNV) between isolates (also referred to as single nucleotide polymorphisms [SNP]), nucleotide differences, gene presence or absence throughout the whole genome, gene allele differences, or overall genetic similarity (e.g., *k*-mers, average nucleotide identity, and multiple genome alignment; Konstantinidis et al., 2006; Sims et al., 2009; Nielsen et al., 2011; Maiden et al., 2013; Leekitcharoenphon et al., 2014). Once sequenced using a next-generation sequencer such as the IonTorrentTM (Thermo Fisher Scientific, Waltham, USA)¹ or HiSeqTM/MiSeqTM (Illumina Inc., San Diego, USA)² platforms, one or two files are generated depending on if single-end or paired sequencing is done, respectively. The file(s) contain all the “raw” sequence reads of the genomic fragments of ~150–300 bp in length, which can then be analyzed directly or assembled into contigs to form a draft genome. The assembly can be done with the help of a reference genome (e.g., *E. coli* O157:H7

¹Thermo Fisher Scientific, Waltham, USA IonTorrentTM, Thermo Fisher Scientific. Available online at: <https://www.thermofisher.com/ca/en/home/brands/ion-torrent.html> (Accessed January 12, 2016).

²Illumina Inc., San Diego, USA Illumina Sequencing Platform Systems Portfolio. Available online at: <http://www.illumina.com/content/dam/illumina-marketing/npassets/PDF/brochure-sequencing-systems-portfolio.pdf> (Accessed January 12, 2016).

str. Sakai) or *de novo*. Most methods used to demonstrate the utility of WGS in public health investigation of bacteria have used assembly-based analysis. In general, assembly uses one of more than 10 assemblers available, but microbiological investigations, especially for *Enterobacteriaceae*, have generally used Velvet or Burrows-Wheeler Aligner (Zerbino and Birney, 2008; Li and Durbin, 2009). Once assembled, the genome can then be compared to other genomes to identify similarity. This is the step in which there are the diverse aforementioned analysis methods with each having multiple different software algorithms and parameters in which to approach them. These approaches are bundled into analysis “pipelines” in which a raw or assembled genome sequence file can be inputted and subjected to multiple software algorithms with specified settings (Kisand and Lettieri, 2013).

Most genome analysis pipelines require high performance computing. There are, however, commercial methods emerging that have optimized genome assembly and analysis to run on high-performance desktop computers or utilize external computing infrastructure to run the computationally intensive steps. These commercial methods include: Bionumerics (Applied Maths NV, Sint-Martens-Latem, Belgium)³, CLC Genomic Workbench (Qiagen, Redwood City, CA, USA) and RidomTM SeqSphere+ (Ridom GmbH, Münster, Germany)⁴. For SNV analyses, there are also online web-interfaces that allow the user to upload a set of sequence files and use a genomic center’s computing infrastructure to run the software (e.g., SeqSero [http://www.denglab.info/SeqSero] and the Center for Genomic Epidemiology [http://www.genomicepidemiology.org]).

WGS IN STEC SURVEILLANCE

SNV analysis has been the predominant method used to date to type isolates using WGS. Usually only the portion of the genome that is conserved amongst a species or a specific pathovar is used for determining strain relatedness (Tettelin et al., 2005; Maiden et al., 2013). This type of analysis is coined “core SNV” analysis to differentiate from SNV analysis of the entire genome. Four groups have applied SNV analysis for the typing of STEC for clinical public health purposes: ProVLab in collaboration with PulseNet Canada and the Public Health Agency of Canada National Microbiology Laboratory (PHAC-NML), the Danish Center for Genomic Epidemiology, Health Protection Scotland and Public Health England.

The Serum Staten Institute in Copenhagen, Denmark sequenced 42 isolates received in a 7-week period and determined their relationship using their web-based tools SNPtree and NDtree (Joensen et al., 2014). During this study period they had an outbreak with 13 cases of *E. coli* O157:H7, six of which were included in their study. The NDtree method, an assembly-free approach that compares the test isolates to nucleotide segments of a reference genome and generates a score representing the

differences in nucleotides found between the genomes, was able to distinguish the outbreak O157 isolates from the other six O157 non-outbreak isolates and the other STEC serotypes. The SNPtree method clustered all serotypes except for O117 K1:H7 together and found 29–65 SNVs different within the outbreak O157 isolates and 521–753 SNVs different between the outbreak and non-outbreak O157 isolates. This group also demonstrated the ability of SNPtree and NDtree methods to differentiate *Salmonella* Typhimurium strains from each other (Leekitcharoenphon et al., 2014).

Public Health England has demonstrated that SNV phylogenetic methods can accurately identify outbreak isolates while adding increased sensitivity to current methods (MLVA and epidemiological investigations in this case; Dallman et al., 2015a). In one of their studies, 572 isolates received by the Gastrointestinal Bacterial Reference Unit for typing including randomly selected isolates from 2012 ($n = 334$) and 2013 ($n = 147$) were sequenced. Based on temporal and epidemiological linkages, the maximum number of SNV differences for isolates to be part of the same cluster was found to be five. An intriguing part of this study was that SNV analysis identified two outbreaks that were not detected by their routine epidemiological investigations or MLVA typing, but were later found to have previously unrecognized epidemiological linkages. This group also demonstrated that SNV analysis was concordant with epidemiological investigations in its ability to identify two different outbreaks caused by watercress contaminated with *E. coli* O157 from two different retailers (one supplied by imports from North America and Europe and the other supplied from south England; Jenkins et al., 2015). SNV analysis may also be applicable to non-O157 because in one study of a nursery school-associated *E. coli* O26:H11 outbreak, ≤ 3 SNVs differences were found in outbreak associated-isolates compared to ≥ 272 SNVs differences between outbreak and non-outbreak isolates (Dallman et al., 2015b).

In Scotland, using SNV methods, a 5-year retrospective review of 105 *E. coli* O157 isolates and 11 epidemiologically linked clusters, found that WGS was generally concordant with MLVA (Holmes et al., 2015). In this study, epidemiologically linked cases exhibited SNV differences of ≤ 4 , while unrelated cases had SNV differences between 9 and 1632. Two sets of isolates that differed in only one MLVA locus were 32 and 126 SNVs different from the other isolate in each set, demonstrating the increased discriminatory power of WGS.

In 2014, Alberta had one of the largest outbreaks of *E. coli* O157:H7 since monitoring by PulseNet Canada began in 2000 with a final tally of 119 clinical cases, which was linked to pork consumption (ProMed-mail post 2759887, 2014). The PHAC-NML SNVPhyl pipeline was used to detect SNV in 111 of these clinical cases and 6 environmental/food isolates and was compared to the current protocol for identifying outbreaks, which involves PFGE and MLVA profiling in combination with epidemiological investigations (Sabat et al., 2013). Clinical, food and environmental isolates from the pork-associated outbreak were found to have ≤ 23 SNVs different from each other and a minimum of 84 SNVs different from isolates not associated with the outbreak, which included sporadic isolates, a concurrent

³Applied Maths NV, Sint-Martens-Latem, Belgium BioNumerics Seven: a unique software platform | Applied Maths. Available online at: <http://www.applied-maths.com/bionumerics> (Accessed January 12, 2016).

⁴Ridom GmbH, Münster, Germany Ridom SeqSphere+. Available online at: http://www.ridom.de/seqsphere/u/User_Guide.html (Accessed January 12, 2016).

smaller outbreak associated with a summer fair and a 2012 beef outbreak. The intra-outbreak SNV differences in the two other outbreaks were 0–5 SNVs and it should be noted that 109 of the 117 sequenced pork outbreak isolates were also 0–5 SNVs different from each other. The same isolates were also subjected to *k*-mer analysis in which the genome is segmented into nucleotide sequences of a pre-determined length (in this case 25-mers) and the frequency of each *k*-mer in the entire genome of each isolate is compared to the frequency of *k*-mers in all other isolates to determine a *k*-mer phylogeny tree. This analysis was able to cluster each outbreak into separate nodes. Interestingly, the *k*-mer method was also able to distinguish isolates from within the pork-associated outbreak that lacked some virulence genes present in all other outbreak isolates. The current surveillance methods and SNV analysis could not distinguish the isolates missing virulence genes from other isolates of the same outbreak. The difference between the two WGS methods is likely because SNV analysis compares conserved genomic regions where virulence factors are rarely found, whereas *k*-mer analysis analyzes the entire genome.

WGS ANALYSIS APPROACHES

The ideal bacterial typing method should have the following characteristics: accuracy, inter- and intra-laboratory reproducibility, stability with multiple passaging of isolates, high discriminatory power, concordance with epidemiological data, speed, ease-of-use, cost effectiveness, and amenability to computerized analysis (Van Belkum et al., 2007). WGS fulfills many of these requirements while providing better accuracy and discriminatory power than PFGE, MLVA and/or epidemiological investigations combined. WGS is currently being used by Public Health England for routine pathogen surveillance (Ashton et al., 2015), and PulseNet Canada is currently setting up a similar infrastructure. Before WGS becomes the international standard, especially for networks such as PulseNet, many issues still need to be addressed.

First of all, a comparison of the different software platforms needs to be performed against a set of isolates with known epidemiological and typing data. Although the studies discussed herein demonstrate that different pipelines can cluster outbreak isolates together with similar intra-outbreak SNV differences, the performance of each pipeline should be tested using the same set of isolates. Also, the SNV and *k*-mer analysis methods are very computer intensive (*k*-mer more so than SNV) and are not amenable to use by labs without the appropriate computing infrastructure or expertise. If SNV and/or *k*-mer methods were used as a standard, a new isolate would need to be compared to a curated regional, national and/or international database of isolates to place it into phylogenetic clusters, similar to viral genotyping. Other options such as whole-genome or core genome multi-locus sequence typing (wg or cgMLST), which look at most or all genes in an organism's genome, can create a barcode by assigning numbers for allelic variants and can be run on a desktop computer (Kohl et al., 2014; Leopold et al., 2014; Ruppitsch et al., 2015). Once a

wgMLST database for *E. coli* is developed in PulseNet, it could become a strong contender for the standard WGS typing of STEC.

Once standards and issues surrounding the computing power and expertise are resolved, WGS will supplant current typing methods for STEC and most other organisms. It can also replace serotyping of STEC and typing of *stx* genes (Joensen et al., 2014, 2015) while providing the additional benefits of using the sequence for detecting virulence genes other than *stx* and providing data for research into genetic elements that influence pathogenicity. Finally, if the methods to perform metagenomics on stool are refined, WGS may also replace PCR or selective agars as the initial screening mechanism for STEC and other enteric pathogens. However, there remain significant technical, economic and organizational hurdles to overcome, before the practical use of genome-wide typing and analysis approaches in routine STEC investigations become a reality (Köser et al., 2012; Franz et al., 2014).

EXPERIENCES AT THE ALBERTA PROVLAB

There have been continuous improvements in STEC detection through bacterial culturing, immunochemistry, and molecular and genomic methods. The advancements in each of these methods aim to increase assay sensitivity, specificity, speed, throughput and broad-ranging strain inclusivity. Despite the improvements in commercial assays and technologies, the adoption of detection methods that encompass non-O157 STEC serotypes by frontline laboratories has been slow. For this reason, there is likely to be a continued significant underreporting of STEC infections in current surveillance data in many countries. At the Alberta ProvLab, multiple studies of regional STEC rates and serotypes revealed that a diverse range of serotypes exist among non-O157 STEC in the province; 99% of these were identified by culture method and included O157 ($n = 99$), top six ($n = 45$), and non-top six ($n = 36$) (Chui et al., 2011, 2015a,b; Couturier et al., 2011). Most of the non-O157 STEC identified would not have been detected by routine frontline testing, which is restricted to detect only O157 STEC. Notably, between 8.3 and 52.9% of the non-O157 serotypes identified in these studies were positive for Stx2. Therefore, these findings reveal the underreporting of non-O157 STEC and the potentially pathogenic strains that risk being undetected in the population (Couturier et al., 2011; Chui et al., 2011, 2015a,b; Gould et al., 2013).

CONCLUSION

There has been much dialog surrounding strategies for more comprehensive STEC identification by frontline laboratories. Primarily these approaches focus on the inclusion of Stx typing in routine testing. More recently, the development and incorporation of WGS methods in STEC surveillance aim to improve the epidemiological tracking of infections. As a

secondary benefit of adopting WGS, will be an enhancement in our understanding of STEC biology through the vast collection of genome sequences. Although, there is great promise in WGS in STEC characterization and surveillance, STEC detection will likely continue to rely on a combination of culturing and non-culture methods. As such, regardless of the technologies that arise for STEC detection and characterization, for at least the immediate future, frontline laboratories will continue to need logical testing algorithms that incorporate a selection of the appropriate methods above.

REFERENCES

- Ammon, A., Petersen, L. R., and Karch, H. (1999). A large outbreak of hemolytic uremic syndrome caused by an unusual sorbitol-fermenting strain of *Escherichia coli* O157:H-. *J. Infect. Dis.* 179, 1274–1277. doi: 10.1086/314715
- Ashton, P. M., Nair, S., Peters, T., Bale, J., Powell, D. G., Painset, A., et al. (2015). Identification and typing of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ PrePrints* 3:e1778 doi: 10.7287/peerj.preprints.1425v1
- Bélanger, S. D., Boissinot, M., Ménard, C., Picard, F. J., and Bergeron, M. G. (2002). Rapid detection of Shiga toxin-producing bacteria in feces by multiplex PCR with molecular beacons on the smart cycler. *J. Clin. Microbiol.* 40, 1436–1440. doi: 10.1128/JCM.40.4.1436-1440.2002
- Bettelheim, K. A. (1998). Reliability of CHROMagar O157 for the detection of enterohaemorrhagic *Escherichia coli* (EHEC) O157 but not EHEC belonging to other serogroups. *J. Appl. Microbiol.* 85, 425–428.
- Biolog, Hayward, CA, USA (2008). *Biolog. Rainbow® Agar O157 Technical Information*. Available online at: http://www.biolog.com/docroot_biologco/wwwRoot/pdf/RBO157Tech00P037A.pdf (Accessed October 19, 2015).
- Brian, M. J., Frosolono, M., Murray, B. E., Miranda, A., Lopez, E. L., Gomez, H. F., et al. (1992). Polymerase chain reaction for diagnosis of enterohemorrhagic *Escherichia coli* infection and hemolytic-uremic syndrome. *J. Clin. Microbiol.* 30, 1801–1806.
- Buss, S. N., Leber, A., Chapin, K., Fey, P. D., Bankowski, M. J., Jones, M. K., et al. (2015). Multicenter Evaluation of the BioFire FilmArray™ Gastrointestinal panel for the etiologic diagnosis of infectious gastroenteritis. *J. Clin. Microbiol.* 53, 915–925. doi: 10.1128/JCM.02674-14
- Centers for Disease Control and Prevention (CDC) (2001). University outbreak of calicivirus infection mistakenly attributed to Shiga toxin-producing *Escherichia coli* O157:H7–Virginia, 2000. *MMWR Morb. Mortal. Wkly. Rep.* 50, 489–491. Available online at: www.cdc.gov/mmwr/preview/mmwrhtml/mm5023a3.htm
- Centers for Disease Control and Prevention (CDC) (2006). Importance of culture confirmation of Shiga toxin-producing *Escherichia coli* infection as illustrated by outbreaks of gastroenteritis–New York and North Carolina, 2005. *MMWR Morb. Mortal. Wkly. Rep.* 55, 1042–1045. Available online at: www.cdc.gov/mmwr/preview/mmwrhtml/mm5538a3.htm
- Chattaway, M. A., Dallman, T. J., Gentle, A., Wright, M. J., Long, S. E., Ashton, P. M., et al. (2016). Whole genome sequencing for public health surveillance of Shiga Toxin-Producing *Escherichia coli* other than serogroup O157. *Front. Microbiol.* 3:258. doi: 10.3389/fmicb.2016.00258
- Chou, F., Croxen, M. A., Jassem, A. N., Pintar, K., Pollari, F., Paccagnella, A., et al. (2014). “Pooled nucleic acid amplification test for screening of stool specimens for shiga toxin-producing *Escherichia coli*, in *The Canadian Journal of Infectious Diseases & Medical Microbiology* (Victoria BC: Pulsus Group Inc.), e39.
- CHROMagar™ O157, CHROMagar Microbiology, and Paris, France, V. (2013). *CHROMagar™ O157, Instructions for Use*. Available online at: http://www.chromagar.com/products-chromagar-o157-focus-on-e-coli-o157-26.html#_VpUvf5MrLdQ (Accessed January 12, 2016).
- CHROMagar™ STEC CHROMagar Microbiology, and Paris, France (2014). *CHROMagar™ STEC. Instructions for Use*. Available online at: http://www.chromagar.com/fichiers/1407415133NT_EXT_058_V7_ST.pdf (Accessed January 12, 2016).
- Chui, L., Couturier, M. R., Chiu, T., Wang, G., Olson, A. B., McDonald, R. R., et al. (2010). Comparison of Shiga toxin-producing *Escherichia coli* detection methods using clinical stool samples. *J. Mol. Diagn.* 12, 469–475. doi: 10.2353/jmoldx.2010.090221
- Chui, L., Lee, M. C., Allen, R., Bryks, A., Haines, L., and Boras, V. (2013). Comparison between ImmunoCard STAT!® and real-time PCR as screening tools for both O157:H7 and non-O157 Shiga toxin-producing *Escherichia coli* in Southern Alberta, Canada. *Diagn. Microbiol. Infect. Dis.* 77, 8–13. doi: 10.1016/j.diagmicrobio.2013.05.015
- Chui, L., Lee, M. C., Malejczyk, K., Lim, L., Fok, D., and Kwong, P. (2011). Prevalence of shiga toxin-producing *Escherichia coli* as detected by enzyme-linked immunoassays and real-time PCR during the summer months in northern Alberta, Canada. *J. Clin. Microbiol.* 49, 4307–4310. doi: 10.1128/JCM.05211-11
- Chui, L., Li, V., Fach, P., Delannoy, S., Malejczyk, K., Patterson-Fortin, L., et al. (2015a). Molecular profiling of *Escherichia coli* O157:H7 and non-O157 strains isolated from humans and cattle in Alberta, Canada. *J. Clin. Microbiol.* 53, 986–990. doi: 10.1128/JCM.03321-14
- Chui, L., Patterson-Fortin, L., Kuo, J., Li, V., and Boras, V. (2015b). Evaluation of enzyme immunoassays and real-time PCR for detecting Shiga toxin-producing *Escherichia coli* in Southern Alberta, Canada. *J. Clin. Microbiol.* 53, 1019–1023. doi: 10.1128/JCM.03288-14
- Church, D. L., Emshey, D., Semeniuk, H., Lloyd, T., and Pitout, J. D. (2007). Evaluation of BBL CHROMagar O157 versus sorbitol-MacConkey medium for routine detection of *Escherichia coli* O157 in a centralized regional clinical microbiology laboratory. *J. Clin. Microbiol.* 45, 3098–3100. doi: 10.1128/JCM.00426-07
- Claas, E. C., Burnham, C. A., Mazzulli, T., Templeton, K., and Topin, F. (2013). Performance of the xTAG® gastrointestinal pathogen panel, a multiplex molecular assay for simultaneous detection of bacterial, viral, and parasitic causes of infectious gastroenteritis. *J. Microbiol. Biotechnol.* 23, 1041–1045. doi: 10.4014/jmb.1212.12042
- Coupland, L. J., McElarney, I., Meader, E., Cowley, K., Alcock, L., Naunton, J., et al. (2013). Simultaneous detection of viral and bacterial enteric pathogens using the Seeplex® Diarrhea ACE detection system. *Epidemiol. Infect.* 141, 2111–2121. doi: 10.1017/S0950268812002622
- Couturier, M. R., Lee, B., Zelyas, N., and Chui, L. (2011). Shiga-toxigenic *Escherichia coli* detection in stool samples screened for viral gastroenteritis in Alberta, Canada. *J. Clin. Microbiol.* 49, 574–578. doi: 10.1128/JCM.01693-10
- Dallman, T. J., Byrne, L., Ashton, P. M., Cowley, L. A., Perry, N. T., Adak, G., et al. (2015a). Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin. Infect. Dis.* 61, 305–312. doi: 10.1093/cid/civ318
- Dallman, T. J., Byrne, L., Launders, N., Glen, K., Grant, K. A., and Jenkins, C. (2015b). The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. *Epidemiol. Infect.* 143, 1672–1680. doi: 10.1017/S0950268814002696
- Deng, J., Luo, X., Wang, R., Jiang, L., Ding, X., Hao, W., et al. (2015). A comparison of Luminex xTAG® Gastrointestinal Pathogen Panel (xTAG GPP) and routine tests for the detection of enteropathogens circulating in Southern China. *Diagn. Microbiol. Infect. Dis.* 83, 325–330. doi: 10.1016/j.diagmicrobio.2015.07.024

AUTHOR CONTRIBUTIONS

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

FUNDING

Salary support for BP was provided through Alberta Innovates - Health Solutions (AIHS) funding of The Alberta Provincial Pediatric EnTeric Infection Team.

- EFSA BIOHAZ, Panel (2013). Scientific Opinion on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment. *EFSA J.* 3138, 1–106. doi: 10.2903/j.efsa.2013.3138
- Fey, P. D., Wickert, R. S., Rupp, M. E., Safranek, T. J., and Hinrichs, S. H. (2000). Prevalence of non-O157:H7 shiga toxin-producing *Escherichia coli* in diarrheal stool samples from Nebraska. *Emerging Infect. Dis.* 6, 530–533. doi: 10.3201/eid0605.000513
- Franz, E., Delaquis, P., Morabito, S., Beutin, L., Gobius, K., Rasko, D. A., et al. (2014). Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int. J. Food Microbiol.* 187, 57–72. doi: 10.1016/j.ijfoodmicro.2014.07.002
- Fratamico, P. M., and Lori, K., Bagi, L. K. (2012). Detection of Shiga toxin-producing *Escherichia coli* in ground beef using the GeneDisc real-time PCR system. *Front. Cell. Infect. Microbiol.* 2:152. doi: 10.3389/fcimb.2012.00152
- Freedman, S. B., Lee, B. E., Louie, M., Pang, X. L., Ali, S., Chuck, A., et al. (2015). Alberta Provincial Pediatric Enteric Infection Team (APPETITE): epidemiology, emerging organisms, and economics. *BMC Pediatr.* 15, 89. doi: 10.1186/s12887-015-0407-7
- Gerritzen, A., Wittke, J. W., and Wolff, D. (2011). Rapid and sensitive detection of Shiga toxin-producing *Escherichia coli* directly from stool samples by real-time PCR in comparison to culture, enzyme immunoassay and Vero cell cytotoxicity assay. *Clin. Lab.* 57, 993–998.
- Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A., and Hewlett, E. L. (2015). Whole-genome sequencing in outbreak analysis. *Clin. Microbiol. Rev.* 28, 541–563. doi: 10.1128/CMR.00075-13
- Gouali, M., Ruckly, C., Carle, I., Lejay-Collin, M., and Weill, F. X. (2013). Evaluation of CHROMagar STEC and STEC O104 chromogenic agar media for detection of Shiga Toxin-producing *Escherichia coli* in stool specimens. *J. Clin. Microbiol.* 51, 894–900. doi: 10.1128/JCM.03121-12
- Gould, L. H., Bopp, C., Strockbine, N., Atkinson, R., Baselski, V., Body, B., et al. (2009). Recommendations for diagnosis of shiga toxin-producing *Escherichia coli* infections by clinical laboratories. *MMWR Recomm. Rep.* 58, 1–14. Available online at: www.cdc.gov/mmwr/preview/mmwrhtml/rr5812a1.htm
- Gould, L. H., Mody, R. K., Ong, K. L., Clogher, P., Cronquist, A. B., Garman, K. N., et al. (2013). Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000–2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Pathog. Dis.* 10, 453–460. doi: 10.1089/fpd.2012.1401
- Grad, Y. H., Lipsitch, M., Feldgarden, M., Arachchi, H. M., Cerqueira, G. C., Fitzgerald, M., et al. (2012). Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3065–3070. doi: 10.1073/pnas.1121491109
- Grif, K., Orth, D., Dierich, M. P., and Würzner, R. (2007). Comparison of an immunochromatographic rapid test with enzyme-linked immunosorbent assay and polymerase chain reaction for the detection of Shiga toxins from human stool samples. *Diagn. Microbiol. Infect. Dis.* 59, 97–99. doi: 10.1016/j.diagmicrobio.2007.03.022
- Grys, T. E., Sloan, L. M., Rosenblatt, J. E., and Patel, R. (2009). Rapid and sensitive detection of Shiga toxin-producing *Escherichia coli* from nonenriched stool specimens by real-time PCR in comparison to enzyme immunoassay and culture. *J. Clin. Microbiol.* 47, 2008–2012. doi: 10.1128/JCM.02013-08
- Gunzer, F., Böhm, H., Rüssmann, H., Bitzan, M., Aleksic, S., and Karch, H. (1992). Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. *J. Clin. Microbiol.* 30, 1807–1810.
- Hara-Kudo, Y., Nemoto, J., Ohtsuka, K., Segawa, Y., Takatori, K., Kojima, T., et al. (2007). Sensitive and rapid detection of Vero toxin-producing *Escherichia coli* using loop-mediated isothermal amplification. *J. Med. Microbiol.* 56, 398–406. doi: 10.1099/jmm.0.46819-0
- Hermos, C. R., Janineh, M., Han, L. L., and McAdam, A. J. (2011). Shiga toxin-producing *Escherichia coli* in children: diagnosis and clinical manifestations of O157:H7 and non-O157:H7 infection. *J. Clin. Microbiol.* 49, 955–959. doi: 10.1128/JCM.02119-10
- Hirvonen, J. J., Siitonen, A., and Kaukoranta, S. S. (2012). Usability and performance of CHROMagar STEC medium in detection of Shiga toxin-producing *Escherichia coli* strains. *J. Clin. Microbiol.* 50, 3586–3590. doi: 10.1128/JCM.01754-12
- Holmes, A., Allison, L., Ward, M., Dallman, T. J., Clark, R., Fawkes, A., et al. (2015). The utility of Whole Genome Sequencing of *Escherichia coli* O157 for outbreak detection and epidemiological surveillance. *J. Clin. Microbiol.* 53, 3565–73. doi: 10.1128/JCM.01066-15
- Jelacic, J. K., Damrow, T., Chen, G. S., Jelacic, S., Bielaszewska, M., Ciol, M., et al. (2003). Shiga toxin-producing *Escherichia coli* in Montana: bacterial genotypes and clinical profiles. *J. Infect. Dis.* 188, 719–729. doi: 10.1086/376999
- Jenkins, C., Dallman, T. J., Launders, N., Willis, C., Byrne, L., Jorgensen, F., et al. (2015). Public health investigation of two outbreaks of Shiga Toxin-Producing *Escherichia coli* O157 associated with consumption of watercress. *Appl. Environ. Microbiol.* 81, 3946–3952. doi: 10.1128/AEM.04188-14
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Karmali, M. A., Petric, M., Lim, C., Cheung, R., and Arbus, G. S. (1985). Sensitive method for detecting low numbers of verotoxin-producing *Escherichia coli* in mixed cultures by use of colony sweeps and polymyxin extraction of verotoxin. *J. Clin. Microbiol.* 22, 614–619.
- Kase, J. A., Maounoun-Lasari, A., Son, I., Lin, A., and Hammack, T. S. (2015). Comparison of eight different agars for the recovery of clinically relevant non-O157 Shiga toxin-producing *Escherichia coli* from baby spinach, cilantro, alfalfa sprouts and raw milk. *Food Microbiol.* 46, 280–287. doi: 10.1016/j.fm.2014.08.020
- Khare, R., Espy, M. J., Cebelski, E., Boxrud, D., Sloan, L. M., Cunningham, S. A., et al. (2014). Multiplex detection of gastrointestinal pathogens: a comparative evaluation of two commercial panels using clinical stool specimens. *J. Clin. Microbiol.* 52, 3667–3673. doi: 10.1128/JCM.01637-14
- Kisand, V., and Lettieri, T. (2013). Genome sequencing of bacteria: sequencing, de novo assembly and rapid analysis using open source tools. *BMC Genomics* 14:211. doi: 10.1186/1471-2164-14-211
- Kohl, T. A., Diel, R., Harmsen, D., Rothgänger, J., Walter, K. M., Merker, M., et al. (2014). Whole-genome-based Mycobacterium tuberculosis surveillance: a standardized, portable, and expandable approach. *J. Clin. Microbiol.* 52, 2479–2486. doi: 10.1128/JCM.00567-14
- Konowalchuk, J., Speirs, J. I., and Stavric, S. (1977). Vero response to a cytotoxin of *Escherichia coli*. *Infect. Immun.* 18, 775–779.
- Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006). Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl. Environ. Microbiol.* 72, 7286–7293. doi: 10.1128/AEM.01398-06
- Köser, C. U., Ellington, M. J., Cartwright, E. J., Gillespie, S. H., Brown, N. M., Farrington, M., et al. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 8:e1002824. doi: 10.1371/journal.ppat.1002824
- Koziel, M., Kiely, R., Blake, L., O'Callaghan, I., Corcoran, G. D., Lucey, B., et al. (2013). Improved detection of bacterial pathogens in patients presenting with gastroenteritis by use of the EntericBio real-time Gastro Panel I assay. *J. Clin. Microbiol.* 51, 2679–2685. doi: 10.1128/JCM.00809-13
- Leekitcharoenphon, P., Nielsen, E. M., Kaas, R. S., Lund, O., and Aarestrup, F. M. (2014). Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PLoS ONE* 9:e87991. doi: 10.1371/journal.pone.0087991
- Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol.* 52, 2365–2370. doi: 10.1128/JCM.00262-14
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Maiden, M. C. J., Jansen van Rensburg, M. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi: 10.1038/nrmicro3093

- March, S. B., and Ratnam, S. (1986). Sorbitol-MacConkey medium for detection of *Escherichia coli* O157:H7 associated with hemorrhagic colitis. *J. Clin. Microbiol.* 23, 869–872.
- Mariani-Kurkdjian, P., Bingen, E., Gault, G., Silva, N. J., and Weill, F. (2011). *Escherichia coli* O104:H4 south-west France, June 2011. *Lancet Infect. Dis.* 11, 732–733. doi: 10.1016/S1473-3099(11)70266-3
- McCallum, C., McGregor, A., and Vanniasinkam, T. (2013). Prevalence of Shiga toxin-producing *Escherichia coli* (STEC) in Tasmania, Australia. *Pathology* 45, 681–688. doi: 10.1097/PAT.0000000000000000
- Mengelle, C., Mansuy, J. M., Prere, M. F., Grouteau, E., Claudet, I., Kamar, N., et al. (2013). Simultaneous detection of gastrointestinal pathogens with a multiplex Luminex-based molecular assay in stool samples from diarrhoeic patients. *Clin. Microbiol. Infect.* 19, E458–E465. doi: 10.1111/1469-0691.12255
- Milley, D. G., and Sekla, L. H. (1993). An enzyme-linked immunosorbent assay-based isolation procedure for verotoxigenic *Escherichia coli*. *Appl. Environ. Microbiol.* 59, 4223–4229.
- Mori, Y., and Notomi, T. (2009). Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. *J. Infect. Chemother.* 15, 62–69. doi: 10.1007/s10156-009-0669-9
- Navidad, J. F., Griswold, D. J., Gradus, M. S., and Bhattacharyya, S. (2013). Evaluation of Luminex xTAG gastrointestinal pathogen analyte-specific reagents for high-throughput, simultaneous detection of bacteria, viruses, and parasites of clinical and public health importance. *J. Clin. Microbiol.* 51, 3018–3024. doi: 10.1128/JCM.00896-13
- Ngwa, G. A., Schop, R., Weir, S., León-Velarde, C. G., and Odumeru, J. A. (2013). Detection and enumeration of *E. coli* O157:H7 in water samples by culture and molecular methods. *J. Microbiol. Methods* 92, 164–172. doi: 10.1016/j.mimet.2012.11.018
- Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28:e63. doi: 10.1093/nar/28.12.e63
- Onori, M., Coltella, L., Mancinelli, L., Argentieri, M., Menichella, D., Villani, A., et al. (2014). Evaluation of a multiplex PCR assay for simultaneous detection of bacterial and viral enteropathogens in stool samples of paediatric patients. *Diagn. Microbiol. Infect. Dis.* 79, 149–154. doi: 10.1016/j.diagmicrobio.2014.02.004
- Pai, C. H., Gordon, R., Sims, H. V., and Bryan, L. E. (1984). Sporadic cases of hemorrhagic colitis associated with *Escherichia coli* O157:H7. Clinical, epidemiologic, and bacteriologic features. *Ann. Intern. Med.* 101, 738–742.
- Perera, L. P., Marques, L. R., and O'Brien, A. D. (1988). Isolation and characterization of monoclonal antibodies to Shiga-like toxin II of enterohemorrhagic *Escherichia coli* and use of the monoclonal antibodies in a colony enzyme-linked immunosorbent assay. *J. Clin. Microbiol.* 26, 2127–2131.
- Pollard, D. R., Johnson, W. M., Lior, H., Tyler, S. D., and Rozee, K. R. (1990). Rapid and specific detection of verotoxin genes in *Escherichia coli* by the polymerase chain reaction. *J. Clin. Microbiol.* 28, 540–545.
- ProMed-mail post 2759887 (2014). ProMED-mail post. Available online at: <http://www.promedmail.org/direct.php?id=2759887> (Accessed October 19, 2015).
- Radu, S., Rusul, G., Ling, O. W., Purwati, E., Mustakim, M., and Lihan, S. (2000). Rapid isolation and detection of *Escherichia coli* O157:H7 by use of rainbow agar O157 and PCR assay. *Southeast Asian J. Trop. Med. Public Health* 31, 77–79. Available online at: <http://imsear.hellis.org/handle/123456789/32177>
- Remis, R. S. (1984). Sporadic cases of hemorrhagic colitis associated with *Escherichia coli* O157:H7. *Ann. Intern. Med.* 101:624. doi:10.7326/0003-4819-101-5-624
- Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H. L., Allerberger, F., et al. (2015). Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Listeria monocytogenes*. *J. Clin. Microbiol.* 53, 2869–2876. doi: 10.1128/JCM.01193-15
- Sabat, A. J., Budimir, A., Nashev, D., Sá-Leão, R., van Dijk, J. M., Laurent, F., et al. (2013). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill.* 18, 20380. Available online: <http://www.eurosurveillance.org/images/dynamic/EE/V18N04/art20380.pdf>
- Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., et al. (2011). Foodborne illness acquired in the United States—major pathogens. *Emerging Infect. Dis.* 17, 7–15. doi: 10.3201/eid1701.091101p1
- Schimmer, B., Nygard, K., Eriksen, H. M., Lassen, J., Lindstedt, B. A., Brandal, L. T., et al. (2008). Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive *Escherichia coli* O103:H25 traced to cured mutton sausages. *BMC Infect. Dis.* 8:41. doi: 10.1186/1471-2334-8-41
- Scotland, S. M., Rowe, B., Smith, H. R., Willshaw, G. A., and Gross, R. J. (1988). Vero cytotoxin-producing strains of *Escherichia coli* from children with haemolytic uraemic syndrome and their detection by specific DNA probes. *J. Med. Microbiol.* 25, 237–243. doi: 10.1099/00222615-25-4-237
- Sims, G. E., Jun, S. R., Wu, G. A., and Kim, S. H. (2009). Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2677–2682. doi: 10.1073/pnas.0813249106
- Smith, K. E., Wilker, P. R., Reiter, P. L., Hedican, E. B., Bender, J. B., and Hedberg, C. W. (2012). Antibiotic treatment of *Escherichia coli* O157 infection and the risk of hemolytic uraemic syndrome, Minnesota. *Pediatr. Infect. Dis. J.* 31, 37–41. doi: 10.1097/INF.0b013e31823096a8
- Soon, J. M., Seaman, P., and Baines, R. N. (2013). *Escherichia coli* O104:H4 outbreak from sprouted seeds. *Int. J. Hyg. Environ. Health* 216, 346–354. doi: 10.1016/j.ijheh.2012.07.005
- STEC, National Surveillance Summary, I. (2014). *Centers for Disease Control and Prevention (CDC). Shiga Toxin-Producing Escherichia coli (STEC) National Surveillance Annual Summary, 2012*. Atlanta: US Department of Health and Human Services, CDC.
- Teel, L. D., Daly, J. A., Jerris, R. C., Maul, D., Svanas, G., O'Brien, A. D., et al. (2007). Rapid detection of Shiga toxin-producing *Escherichia coli* by optical immunoassay. *J. Clin. Microbiol.* 45, 3377–3380. doi: 10.1128/JCM.00837-07
- Teh, C. S., Chua, K. H., Lim, Y. A., Lee, S. C., and Thong, K. L. (2014). Loop-mediated isothermal amplification assay for detection of generic and verocytotoxin-producing *Escherichia coli* among indigenous individuals in Malaysia. *Sci. World J.* 2014:457839. doi: 10.1155/2014/457839
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102
- Thompson, L. H., Giercke, S., Beaudoin, C., Woodward, D., and Wylie, J. L. (2005). Enhanced surveillance of non-O157 verotoxin-producing *Escherichia coli* in human stool samples from Manitoba. *Can. J. Infect. Dis. Med. Microbiol.* 16, 329–334. doi: 10.1155/2005/859289
- Tillman, G. E., Wasilenko, J. L., Simmons, M., Lauze, T. A., Minicozzi, J., Oakley, B. B., et al. (2012). Isolation of Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 from ground beef using modified rainbow agar and post-immunomagnetic separation acid treatment. *J. Food Prot.* 75, 1548–1554. doi: 10.4315/0362-028X.JFP-12-110
- Tutenel, A. (2003). Sensitivity of methods for the isolation of *Escherichia coli* O157 from naturally infected bovine faeces. *Vet. Microbiol.* 94, 341–346. doi: 10.1016/S0378-1135(03)00121-4
- Van Belkum, A., Tassios, P. T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N. K., et al. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect.* 13 (Suppl. 3), 1–46. doi: 10.1111/j.1469-0691.2007.01786.x
- Vocale, C., Rimoldi, S. G., Pagani, C., Grande, R., Pedna, F., Arghittu, M., et al. (2015). Comparative evaluation of the new xTAG GPP multiplex assay in the laboratory diagnosis of acute gastroenteritis. Clinical assessment and potential application from a multicentre Italian study. *Int. J. Infect. Dis.* 34, 33–37. doi: 10.1016/j.ijid.2015.02.011
- Wang, F., Jiang, L., and Ge, B. (2012). Loop-mediated isothermal amplification assays for detecting shiga toxin-producing *Escherichia coli* in ground beef and human stools. *J. Clin. Microbiol.* 50, 91–97. doi: 10.1128/JCM.05612-11
- Wasilenko, J. L., Fratamico, P. M., Sommers, C., DeMarco, D. R., Varkey, S., Rhoden, K., et al. (2014). Detection of Shiga toxin-producing *Escherichia coli* (STEC) O157:H7, O26, O45, O103, O111, O121, and O145, and *Salmonella* in retail raw ground beef using the DuPont™ BAX® system. *Front. Cell. Infect. Microbiol.* 4:81. doi: 10.3389/fcimb.2014.00081

- Wells, J. G., Davis, B. R., Wachsmuth, I. K., Riley, L. W., Remis, R. S., Sokolow, R., et al. (1983). Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. *J. Clin. Microbiol.* 18, 512–520.
- Willshaw, G. A., Smith, H. R., Scotland, S. M., Field, A. M., and Rowe, B. (1987). Heterogeneity of *Escherichia coli* phages encoding Vero cytotoxins: comparison of cloned sequences determining VT1 and VT2 and development of specific gene probes. *J. Gen. Microbiol.* 133, 1309–1317. doi: 10.1099/00221287-133-5-1309
- Wong, C. S., Jelacic, S., Habeeb, R. L., Watkins, S. L., and Tarr, P. I. (2000). The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *N. Engl. J. Med.* 342, 1930–1936. doi: 10.1056/NEJM200006293422601
- Wong, C. S., Mooney, J. C., Brandt, J. R., Staples, A. O., Jelacic, S., Boster, D. R., et al. (2012). Risk factors for the hemolytic uremic syndrome in children infected with *Escherichia coli* O157:H7: a multivariable analysis. *Clin. Infect. Dis.* 55, 33–41. doi: 10.1093/cid/cis299
- Wylie, J. L., Van Caesele, P., Gilmour, M. W., Sitter, D., Guttek, C., and Giercke, S. (2013). Evaluation of a new chromogenic agar medium for detection of Shiga toxin-producing *Escherichia coli* (STEC) and relative prevalences of O157 and non-O157 STEC in Manitoba, Canada. *J. Clin. Microbiol.* 51, 466–471. doi: 10.1128/JCM.02329-12
- Yoshitomi, K. J., Jinneman, K. C., Zapata, R., Weagant, S. D., and Fedio, W. M. (2012). Detection and isolation of low levels of *E. coli* O157:H7 in cilantro by real-time PCR, immunomagnetic separation, and cultural methods with and without an acid treatment. *J. Food Sci.* 77, M481–M489. doi: 10.1111/j.1750-3841.2012.02813.x
- Zelyas, N., Poon, A., and Chui, L. (2016). Assessment of commercial chromogenic solid media for the detection of non-O157 Shiga toxin-producing *Escherichia coli* (STEC). *Diag. Microbiol. Infect. Dis.* doi: 10.1016/j.diagmicrobio.2016.03.013. Available online at: <http://www.sciencedirect.com/science/article/pii/S0732889316300530>
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zhang, W., Bielaszewska, M., Bauwens, A., Fruth, A., Mellmann, A., and Karch, H. (2012). Real-time multiplex PCR for detecting Shiga toxin 2-producing *Escherichia coli* O104:H4 in human stools. *J. Clin. Microbiol.* 50, 1752–1754. doi: 10.1128/JCM.06817-11

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

Copyright © 2016 Parsons, Zelyas, Berenger and Chui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



From Farm to Table: Follow-Up of Shiga Toxin-Producing *Escherichia coli* Throughout the Pork Production Chain in Argentina

Rocío Colello, María E. Cáceres, María J. Ruiz, Marcelo Sanz, Analía I. Etcheverría* and Nora L. Padola

OPEN ACCESS

Edited by:

Pina Fratamico,
United States Department
of Agriculture – Agricultural Research
Service, USA

Reviewed by:

James L. Smith,
United States Department
of Agriculture, USA
Yanhong Liu,
European Roma Rights Centre –
United States Department
of Agriculture – Agricultural Research
Service, USA

*Correspondence:

Analía I. Etcheverría
analiain@vet.unicen.edu.ar

Specialty section:

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

Received: 03 November 2015

Accepted: 18 January 2016

Published: 08 February 2016

Citation:

Colello R, Cáceres ME, Ruiz MJ,
Sanz M, Etcheverría AI and Padola NL
(2016) From Farm to Table:
Follow-Up of Shiga Toxin-Producing
Escherichia coli Throughout the Pork
Production Chain in Argentina.
Front. Microbiol. 7:93.
doi: 10.3389/fmicb.2016.00093

Laboratorio de Inmunoquímica y Biotecnología, Centro de Investigación Veterinaria de Tandil – Consejo Nacional de Investigaciones Científicas y Técnicas – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Facultad de Ciencias Veterinarias, Universidad Nacional del Centro de la Provincia de Buenos Aires, Tandil, Argentina

Pigs are important reservoirs of Shiga toxin-producing *Escherichia coli* (STEC). The entrance of these strains into the food chain implies a risk to consumers because of the severity of hemolytic uremic syndrome. This study reports the prevalence and characterization of STEC throughout the pork production chain. From 764 samples, 31 (4.05%) were *stx* positive by PCR screening. At farms, 2.86% of samples were *stx* positive; at slaughter, 4.08% of carcasses were *stx* positive and at boning rooms, 6% of samples were *stx* positive. These percentages decreased in pork meat ready for sale at sales markets (4.59%). From positive samples, 50 isolates could be characterized. At farms 37.5% of the isolates carried *stx1/stx2* genes, 37.5% possessed *stx2e* and 25%, carried only *stx2*. At slaughter we detected 50% of isolates positive for *stx2*, 33% for *stx2e*, and 16% for *stx1/stx2*. At boning rooms 59% of the isolates carried *stx1/stx2*, 14% *stx2e*, and 5% *stx1/stx2/stx2e*. At retail markets 66% of isolates were positive for *stx2*, 17% *stx2e*, and 17% *stx1/stx2*. For the other virulence factors, *ehxA* and *saa* were not detected and *eae* gene was detected in 12% of the isolates. Concerning putative adhesins, *agn43* was detected in 72%, *ehaA* in 26%, *aida* in 8%, and *iha* in 6% of isolates. The strains were typed into 14 *E. coli* O groups (O1, O2, O8, O15, O20, O35, O69, O78, O91, O121, O138, O142, O157, O180) and 10 H groups (H9, H10, H16, H21, H26, H29, H30, H32, H45, H46). This study reports the prevalence and characterization of STEC strains through the chain pork suggesting the vertical transmission. STEC contamination originates in the farms and is transferred from pigs to carcasses in the slaughter process and increase in meat pork at boning rooms and sales markets. These results highlight the need to implement an integrated STEC control system based on good management practices on the farm and critical control point systems in the food chain.

Keywords: STEC, foodborne pathogens, pork production chain, prevalence, characterization

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are important foodborne pathogens that can cause severe disease, including a life-threatening complication such as bloody diarrhea and hemolytic uremic syndrome (HUS; Paton and Paton, 1998). HUS is one of the most common etiologies for acute kidney injury and an important cause of acquired chronic kidney disease in children (Grisaru, 2014). This damage is produced by the action of cytotoxins Stx1 and Stx2, being Stx2 and their subtypes associated more frequently with HUS (Beutin et al., 2007). The ability to adhere to epithelial cells is an important virulence trait, because adherence presumably enables to deliver toxins efficiently to host organs (Tarr et al., 2000). Intimin, encoded by *eae* gene, is required for intimate bacterial adhesion to epithelial cells inducing a characteristic histopathological lesion defined as “attaching and effacing” (A/E) and has been considered as a risk factor for disease in human (Ethelberg et al., 2004). However, the presence of *eae* would not be essential for pathogenesis, considering that some *eae* negative STEC have been associated with severe disease in human (Paton et al., 2001). Some studies reported adherence factors other than intimin, such as Saa (Paton et al., 2001), AIDA and Agn43 (Restieri et al., 2007), EhaA (Wu et al., 2010), Iha (Tarr et al., 2000; Szalo et al., 2002). AIDA was identified in diffusely adhering diarrheagenic *E. coli* strain and is associated with edema disease and diarrhea in pigs (Niewerth et al., 2001), contributing to bacterial intercellular aggregation and biofilm formation (Restieri et al., 2007); *iha* encode for an outer membrane protein identified as a bacterial adherence conferring iron regulated gene (Tarr et al., 2000) and Agn43 and EhaA are autotransporter proteins of O157:H7 involved in adhesion and biofilm formation (Wells et al., 2008). Other factors are also involved in human pathogenicity such as a plasmid that encoded enterohemolysin (EhxA), among others (Feng and Reddy, 2013).

Argentina, where the HUS is endemic, hold the highest record worldwide of this syndrome with an incidence of 17/100,000 children less than 5 years old (Rivas et al., 2010). Although STEC O157:H7 is recognized as the most important serotype associated with human infection, there are more than 400 non-O157 serotypes that have been involved in human disease and isolated from different reservoirs including cattle, pigs, goats, sheep, cats, and dogs (Parma et al., 2000; Padola et al., 2004; Amézquita-López et al., 2014). STEC usually do not produce disease in animals, however, the Stx2e subtype is involved in edema disease in pigs, a peracute toxemia characterized by vascular necrosis, edema, neurological signs and that in some cases can be fatal (Niewerth et al., 2001). STEC strains have been isolated from pork products and have been associated with human infections as diarrhea and HUS, including strains harboring *stx2e* subtype (Sonntag et al., 2005; Kaufmann et al., 2006; Trotz-Williams et al., 2012); however, it is unknown if the contamination of pork-derivate food occurs during the processing or by cross contamination (Tseng et al., 2014).

There is an increase in worldwide demand for fast-growing species with efficient feed conversion rates, such as pigs, because they represent a major share in the growth in the livestock

subsector (Food and Agriculture Organization [FAO], 2014). Because of the limited epidemiologic data of STEC in pork and the increasing role of non-O157 STEC in human illnesses, it is very important to study the role of pigs as reservoirs of STEC and the transmission to the swine production chain (Ercoli et al., 2015). Taking into account the data mentioned above, the aim of this study was to determine the prevalence and to characterize STEC throughout the pork production chain in Argentina.

MATERIALS AND METHODS

Management of Farms and Animals

The study was conducted in two pig production farm systems. Both farms are intensively organized in total confinement. The production stages are: gestation, farrowing, weaning, and growing/finishing (fattening), which are geographically separated from each other within the same farm. The usual group size varies between 10 and 30 pigs. Pigs and employees move from one building to others by corridors that are isolated from external traffic.

Management of Carcasses Until Retails Markets

Pigs at finishing stage are transported to the slaughterhouses. After slaughtered, the pork carcasses are chilled during 24 to 48 h and sent to boning rooms in refrigerated trucks.

At the boning rooms the carcasses were boning to products such as meat and minced meat. Finally, the products are transferred to the retail markets.

Collection of Samples

Seven hundred and sixty four samples were collected from May, 2012 to November, 2014 from two pig production systems.

This study was carried out in accordance with the recommendations of the Animal Welfare Committee from the Veterinary Science Faculty, UNCPBA, Resolution 087/02.

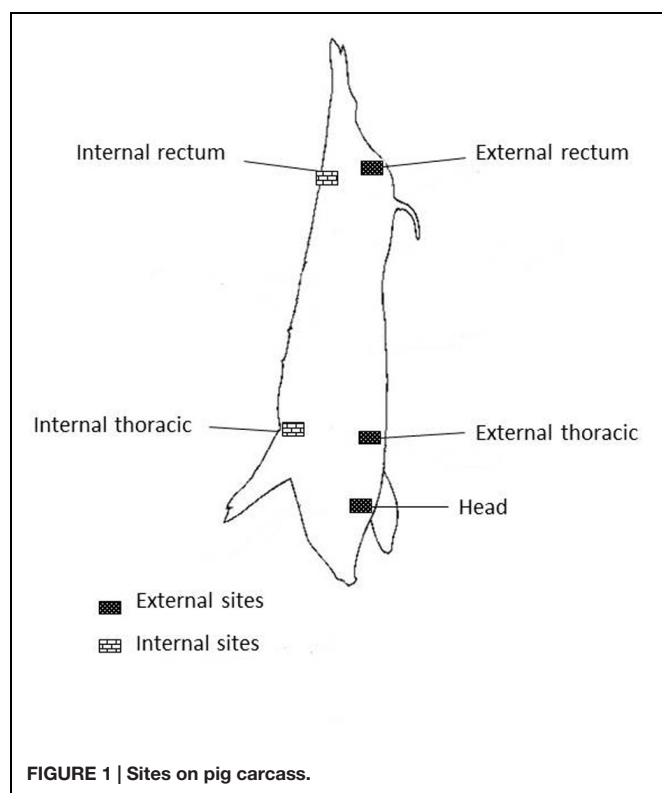
Samples at Farms

Three hundred and forty eight samples were taken at farms. From these, 277 corresponded to rectal swabs, and 71 come from the environment obtained from water drink, feed and feces on the floor by swabbing.

Samples at Slaughterhouses

One hundred and forty seven samples were by swabbing. Of these, 22 were from rectal swabs after slaughter, 85 from carcasses, and 40 from the slaughterhouses environment (pre-washing, scalding, deharing, dressing, cooling, and knives).

Carcasses swabs were taken in concordance with circular No 3496/02 of Servicio Nacional de Sanidad y Calidad Agroalimentaria (SENASA, 2002). Five quarters areas of 100 cm² each one were taken and processed separately, they are named heads (H), external rectum (ER), internal rectum (IR), external thoracic (ET), and internal thoracic (IT; **Figure 1**).



Samples at Boning Rooms

One hundred and eighty one samples were taken. From these, 94 come from carcasses, 24 from meat, 23 from minced meat, and 40 from environmental samples (refrigerated trucks and meat contact surfaces such as meat tables, knives, meat mincing machine, and vertical band saw machine).

Sampling at Retail Markets

Eighty seven samples were taken from retail markets (43 samples come from meat, 13 from minced meat, and 31 from the environment). The environmental samples were obtained from meat tables, knives, vertical band saw machine, and refrigerators.

Sample Preparation and Isolation of STEC

Swabs were processed according to Etcheverría et al. (2010). Briefly, the swabs were cultured in Luria Bertani broth (LB) with shaking at 37°C for 18 h, and then an aliquot was grown on MacConkey agar plates by incubating at 37°C for 24 h.

Ten to fifty individual colonies were processed for amplification of Shiga toxin genes (*stx1*, *stx2*, and *stx2e*; Table 1). Each positive colony for either *stx* was tested for the presence of the *eae*, *ehxA*, and *saa* by multiplex polymerase chain reaction (PCR; Paton and Paton, 2002). Genes encoding adhesins (*ehaA*, *agn43*, *iha*, *aida*) were amplified using monoplex PCR. STEC strains used as positive control were *E. coli* O157:H7 EDL933 (*stx1*, *stx2*, *eae*, *ehxA*, *ehaA*, *agn43*, *iha*), *E. coli* O8 (*stx2e*), *E. coli* O91:H21 (*stx1*, *stx2*, *ehxA*, *saa*), and *E. coli* O157:H19 (*aida*).

Amplification products were separated by electrophoresis on 2% agarose gel containing 0.8 µg/ml of ethidium bromide in running buffer and were visualized by UV transillumination.

Determination of Serotype

O and H types were determined by microagglutination technique in plates and tubes as described by Guinée et al. (1981) and modified by Blanco et al. (1996) using all available O (O1–O175) antisera plus six putative new O antigens (OX176 through OX181) and H (H1–H56) antisera (Pradel et al., 2000).

TABLE 1 | Genes, primers sequences, and size of amplified product of Shiga toxin-producing *Escherichia coli* (STEC).

Gene	Primers sequences (5'–3')	Size of amplified product (bp)	Reference
<i>stx1</i>	stx1F-ATAAATCGCCATTCTGTTGACTA stx1R-AGAACGCCCACTGAGATCATC	180	Paton and Paton, 1998
<i>stx2</i>	stx2F-GGCACTGTCTGAACTGCTCC stx2R-TCGCCAGTTATCTGACATTCTG	255	Paton and Paton, 1998
<i>stx2e</i>	vt23F- CCGTCAGGACTGTCTGAAAC vt24R- GGACGCGATAATTAACCG	663	Woodward et al., 1992
<i>eae</i>	eaeF-GACCCGGCACAAGCATAAGC eaeR-CCACCTGCAGCAACA-AGAGG	384	Paton and Paton, 1998
<i>ehxA</i>	hlyAF-GCATCATCAAGCGTACGTTCC hlyAR-AATGAGCCAAGCTGGTTAAGCT	534	Paton and Paton, 1998
<i>saa</i>	saaDF-CGTGATGAACAGGCTATTGC- saaDR-ATGGACATGCCTGTGGCAAC	119	Paton and Paton, 1998
<i>agn43</i>	agn43F-CTGGAACCGGTCTGCCCTT agn43R-CCTGAACGCCCAAGGGTGATA	433	Restieri et al., 2007
<i>aida</i>	aida1FACAGTATCATATGGAGCCACTC aida1R-TGTGCGCCAGAATAATTAATGG	587	Restieri et al., 2007
<i>ehaA</i>	ehaAF-AGGCATGAGACACGATC ehaAR-d AAGTCGTGCCATTGAGC	500	Wu et al., 2010
<i>iha</i>	ihaF-CAAATGGCTCTTCCGTCATATGC ihaR-CAGGTCGGGGTTACCAAGT	925	Szalo et al., 2002

RESULTS

The results indicate that STEC occurrence is widespread throughout pork production chain. Among the 764 samples, 31 (4.05%) were positive for *stx*. In rectal swabs from the different pig categories, 2.86% were STEC positive, distributed 5.88% at fattening, 4.3% at growing, 2.38% at gestation and 1.51% at farrowing. STEC were not detected in feed, water, and fecal samples taken from farms. At slaughter, 4.08% of carcasses sampled were *stx* positive. The distribution in the different quarters of the carcasses was: 50% from ER, 16.6% from ET, 16.6% from IT, 16.6% from heads. At boning rooms, 6% of samples were STEC positive, belonging 82% to carcasses, and 18% to pork meat. The distribution in the different quarters of the carcasses was: 33.3% from ER, 22.2% from IR, 22.2% from IT, 11.2% from ET and 11% from head. At sale markets 4.59% of STEC positive samples were detected in pork meat ready for sale.

Characterization of STEC

From positive samples, 50 isolates could be characterized by PCR. In samples from farms 6/16 (37.5%) of the isolates carried *stx1/stx2*, 6/16 (37.5%) possessed *stx2e*, and 4/16 (25%) carried *stx2*. At slaughter 3/6 (50%) of isolates were positive for *stx2*, 2/6 (33%) for *stx2e*, and 1/6 (16%) for *stx1/stx2*. At boning rooms 13/22 (59%) of the isolates carried *stx1/stx2*, 3/22 (14%) *stx2e*, and 1/22 (5%) *stx1/stx2/stx2e*. At retail markets 4/6 (66%) of isolates were positive for *stx2*, 1/6 (17%) for *stx2e*, and 1/6 (17%) for *stx1/stx2*. Other virulence factors such as *ehxA* and *saa* were not detected and *eae* was detected in 6/50 (12%) of samples. Concerning putative adhesins, *agn43* was detected in 36/50 (72%), *ehaA* in 13/50 (26%), *aida* in 4/50 (8%), and *iha* in 3/50 (6%) of isolates. The most frequent virulence profiles found were *stx1*, *stx2*, or *stx2e* combined with *agn43* in 36 (72%) strains. The 50 isolates were typed into 14 *E. coli* O groups (O1, O2, O8, O15, O20, O35, O69, O78, O91, O121, O138, O142, O157, O180) and 15 were considered O non-typable (NT). Ten H antigens (H9, H10, H16, H21, H26, H29, H30, H32, H45, H46) were distributed among the 50 strains, while one isolate were non-motile (H-). **Table 2** indicates the relationships between virulence profiles, sites of samples and serotypes in isolated STEC strains.

DISCUSSION

To our knowledge, this study is the first that reports the prevalence and characterization of STEC strains through the chain pork suggesting the vertical transmission of these pathogens. However, there are studies that demonstrate the prevalence in farms, finishing pigs, slaughter, and pork meat in sales markets, separately. The prevalence of STEC in pigs, carcasses, and pork meat at different stage of production from other countries is variable and it is necessary to take caution when comparing prevalence since the variation may be due to several factors, such as sampling method, samples processing, and season in which the study was performed. Different prevalence of STEC in pigs was reported previously, ranged from 2 to 31%, in agreement with our results (Parma et al., 2000; Kaufmann et al.,

2006; Meng et al., 2014). In carcasses at slaughter and boning rooms, the prevalence found is in concordance with other studies whose prevalence ranged from 0.2 to 26% (Leung et al., 2001; Bouvet et al., 2002; Bohaychuk et al., 2011; Koláčková et al., 2014).

At slaughter some operations such as skinning, evisceration and handling are more likely than others to contaminate carcass and meat (Koohmaraie et al., 2005; Etcheverría et al., 2010). For this, some areas of carcasses are more prone than others to be exposure to potential or cross contamination, thus the suggestion of sampling at three or four sites on carcass, because contamination appears to vary considerably among different sites (Roberts et al., 1984). In addition, the ER is the area that involves

TABLE 2 | Relationships between virulence profiles, sites of samples, and serotypes in STEC strains.

Sample site	Serotypes	No. of strains	Virulence profile
Farms (pigs)			
Gestation	O1:HNM	1	<i>stx1</i>
Gestation	O1:HNM	1	<i>stx1, stx2, agn43, ehaA</i>
Fattening	O2:H32	3	<i>stx2, agn43</i>
Growing and Farrowing	O8:H9	4	<i>stx2e, agn43</i>
Fattening	O20:H9	1	<i>stx2, agn43</i>
Fattening	O142:H26	1	<i>stx2</i>
Fattening	ONT:H16	1	<i>stx2e, agn43, iha</i>
Fattening	ONT:H26	1	<i>stx1, stx2, agn43</i>
Fattening	ONT:H32	1	<i>stx2e, agn43, ehaA</i>
Growing	ONT:HNM	1	<i>stx2e, agn43</i>
Fattening	ONT:HNM	1	<i>stx1, stx2, agn43</i>
Slaughterhouse			
H	O1:H9	1	<i>stx2e, ehaA</i>
ER	O91:H21	1	<i>stx1, agn43, ehaA, iha</i>
ER	ONT:H29	1	<i>stx1, stx2</i>
ER	ONT:HNM	1	<i>stx2</i>
ET and IT	O8:HNM	2	<i>stx2e, ehaA</i>
Boning rooms			
ET	O8:HNM	1	<i>stx2</i>
ER	O15:H45	2	<i>stx1, stx2, ehaA</i>
IR	O35:H10	1	<i>stx1, stx2, agn43</i>
ET	O69:HNM	1	<i>stx1, stx2, agn43</i>
H	O78:H45	3	<i>stx1, stx2, agn43</i>
H	O78:HNM	2	<i>stx1, stx2, agn43</i>
ET	O121:H21	1	<i>stx2e</i>
ET	O138:H30	1	<i>stx2e</i>
IT and meat	O157:H21	4	<i>stx2, eae, aida, agn43, ehaA</i>
ET	O180:H21	1	<i>stx1, stx2</i>
ET	ONT:H21	1	<i>stx1, stx2</i>
Meat	ONT:H21	2	<i>stx2e, agn43</i>
IR	ONT:H32	1	<i>stx1, stx2, agn43</i>
IT	ONT:HNM	1	<i>stx2, stx2e, eae, agn43</i>
Sale markets			
Meat	O9:H21	1	<i>stx2, agn43</i>
Meat	O9:HNM	2	<i>stx2, agn43</i>
Meat	ONT:H21	1	<i>stx2e, agn43</i>
Meat	ONT:HNM	1	<i>stx2, stx2e, agn43</i>
meat	ONT:HNM	1	<i>stx2, stx2e, eae, agn43</i>

a particular risk of contamination during early stages of dressing as our result in that the ER was the more contaminated area, in concordance with Bouvet et al. (2002). However, other areas sampled as IT and IR shown more contamination probably due to handling at the boning room.

In our study, the prevalence at sale market was less than that informed by Magwedere et al. (2013) in USA (50%), Martin and Beutin (2011) in Germany (14%) and Lee et al. (2009) Korea (15%). This could be due because these studies were performed in sale markets where meat from different origins were sold and cross contamination during handling can occur. In the present study the samples were obtained from sale markets where only meat pork was sold.

From 50 STEC isolates, *stx1/stx2* and *stx2* occurred more frequently than isolates carrying *stx1*. Epidemiologically, *Stx2*-producing strains are more often related with HUS than strains that produce *Stx1* (Paton and Paton, 2002).

Regarding *stx2e*, its prevalence decreased from pigs at farms to pork meat. Although some authors have reported the presence of *stx2e* in STEC strains in human patients on a few occasions, STEC harboring *stx2e* are more likely to cause edema disease in pigs causing economic losses in pig production (Kaufmann et al., 2006). It is necessary to determine a role of these strains in human infection.

The presence of *eae* detected in isolates that harbor too *stx2*, *stx2e*, *agn43* from boning rooms and sale markets implies a high risk for human health (Tseng et al., 2014). The most prevalent adhesin identified among all isolates and involved in adhesion and biofilm formation was *Agn43*, followed by *EhaA* in agreement with Biscola et al. (2011) and Tseng et al. (2014) which detected them in swine and different sources, respectively. In this study, 8% of strain harbored *AIDA*, similar to that found in South Africa (Mohlalole et al., 2013) and China (Zhao et al., 2009). The *iha* was present in few isolates, but this gene has been detected over 70% of the *eae*-negative STEC strains associated with human clinical cases examined in studies in Germany (Hauser et al., 2013) and Argentina (Galli et al., 2010). The high prevalence of LEE negative STEC isolated from pigs in our study emphasizes the need of further work to better define the role that the attachment proteins outside the LEE may play in the adherence to both pork and human epithelial cells.

Although many serotypes isolated in this study have been detected with low incidence in human disease and rarely associated with outbreaks (Friedrich et al., 2002), they have been isolated from pigs, sheep, cattle and food in other countries

(Beutin et al., 1993; Kaufmann et al., 2006; Beutin et al., 2007). At slaughter and boning rooms serogroups associated with human illnesses such as *E. coli* O91, O121, and O157 were detected in agreement with other studies that recovered these serogroups from pig fecal samples (Desrosiers et al., 2001; Friedrich et al., 2002; Karmali et al., 2003; Bielaszewska et al., 2009; Trotz-Williams et al., 2012; Yoon et al., 2013; Tseng et al., 2014).

CONCLUSION

The presented study investigated the tracking of STEC from the farm to table and indicates that the production of meat pork harbored STEC strains. STEC contamination originated in the farms is transferred from pigs to carcasses in the slaughter process and increase in meat pork at boning rooms and sales markets. Besides, the entrance of these strains into the food chain implies a risk to consumers because of severity of the illness they can cause. If STEC is present in any food product, it has the possibility of causing foodborne illness. In addition to public health problem, the presence of strains carrying the *stx2e* gene is a problem for the pig production because they can cause the edema disease causing important economic losses. In spite of the wealth of data available on this important disease, it is necessary to effectively prevent this contamination by educating employees, retailers and consumers on the appropriate handling and storage of meat. Further studies are needed to provide more systematic data in order to fuel the development of novel approaches for control of STEC in foods, including pork meat.

AUTHOR CONTRIBUTIONS

RC conceived, designed, analyzed the experiments, and wrote the manuscript. MR, MS, and MC did some of the experiments. NP and AE designed some of the experiments, analyzed the data, and revised the manuscript.

ACKNOWLEDGMENTS

Authors thank Guillermo Arroyo for their collaboration during sampling collection, María Rosa Ortiz for her technical assistance. This work was supported by PICT 2010-1655, CIC, and SECAT from Argentina.

REFERENCES

- Amézquita-López, B. A., Quiñones, B., Lee, B. G., and Chaidez, C. (2014). Virulence profiling of Shiga toxin-producing *Escherichia coli* recovered from domestic farm animals in Northwestern Mexico. *Front. Cell. Infect. Microbiol.* 4:7. doi: 10.3389/fcimb.2014.00007
- Beutin, L., Geier, D., Steinrück, H., Zimmermann, S., and Scheut, F. (1993). Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. *J. Clin. Microbiol.* 31, 2483–2488.
- Beutin, L., Miko, A., Krause, G., Pries, K., Haby, S., Steege, K., et al. (2007). Identification of human-pathogenic strains of Shiga toxin-producing *Escherichia coli* from food by a combination of serotyping and molecular typing of Shiga toxin genes. *Appl. Environ. Microbiol.* 73, 4769–4775. doi: 10.1128/AEM.00873-07
- Bielaszewska, M., Stoewe, F., Fruth, A., Zhang, W., Prager, R., Brockmeyer, J., et al. (2009). Shiga toxin, cytolethal distending toxin, and hemolysin repertoires in clinical *Escherichia coli* O91 isolates. *J. Clin. Microbiol.* 47, 2061–2066. doi: 10.1128/JCM.00201-09
- Biscola, F. T., Abe, C. M., and Guth, B. E. C. (2011). Determination of adhesin gene sequences in, and biofilm formation by, O157 and non-O157 Shiga toxin-producing *Escherichia coli* strains isolated from different sources. *Appl. Environ. Microbiol.* 77, 2201–2208. doi: 10.1128/AEM.01920-10

- Blanco, J. E., Blanco, M., Blanco, J., Mora, A., Balaguer, L., Mourino, M., et al. (1996). O serogroups, biotypes, and eae genes in *Escherichia coli* strains isolated from diarrheic and healthy rabbits. *J. Clin. Microbiol.* 34, 3101–3107.
- Bohaychuk, V. M., Gensler, G. E., and Barrios, P. R. (2011). Microbiological baseline study of beef and pork carcasses from provincially inspected abattoirs in Alberta, Canada. *Can. Vet. J.* 52, 1095.
- Bouvet, J., Montet, M., Rossel, R., Le Roux, A., Bavai, C., Ray-Gueniot, S., et al. (2002). Prevalence of verotoxin-producing *Escherichia coli* (VTEC) and *E. coli* O157: H7 in French pork. *J. Appl. Microbiol.* 93, 7–14. doi: 10.1046/j.1365-2672.2002.01672.x
- Desrosiers, A., Fairbrother, J. M., Johnson, R. P., Desautels, C., Letellier, A., and Quesy, S. (2001). Phenotypic and genotypic characterization of *Escherichia coli* verotoxin-producing isolates from humans and pigs. *J. Food Prot.* 64, 1904–1911.
- Ercoli, L., Farneti, S., Ranucci, D., Scuota, S., and Branciar, R. (2015). Role of verocytotoxigenic *Escherichia coli* in the swine production chain. *Ital. J. Food Saf.* 4, 51–56. doi: 10.4081/ijfs.2015.5156
- Etcheverría, A., Padola, N., Sanz, M., Polifroni, R., Krüger, A., Passucci, J., et al. (2010). Occurrence of Shiga toxin-producing *E. coli* (STEC) on carcasses and retail beef cuts in the marketing chain of beef in Argentina. *Meat Sci.* 86, 418–421. doi: 10.1016/j.meatsci.2010.05.027
- Ethelberg, S., Olsen, K., Scheut, F., Jensen, C., Schiellerup, P., Enberg, J., et al. (2004). Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg. Infect. Dis.* 10, 842–847. doi: 10.3201/eid1005.030576
- Feng, P. C., and Reddy, S. (2013). Prevalences of Shiga toxin subtypes and selected other virulence factors among Shiga-toxigenic *Escherichia coli* strains isolated from fresh produce. *Appl. Environ. Microbiol.* 79, 6917–6923. doi: 10.1128/AEM.02455-13
- Food and Agriculture Organization [FAO] (2014). *Pigs and Animal Production* [Online]. Available at: <http://www.fao.org/ag/againfo/themes/en/pigs/production.html> [Accessed December 10, 2014]
- Friedrich, A. W., Bielaszewska, M., Zhang, W.-L., Pulz, M., Kuczius, T., Ammon, A., et al. (2002). *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J. Infect. Dis.* 185, 74–84. doi: 10.1086/338115
- Galli, L., Miliwsky, E., Irino, K., Leotta, G., and Rivas, M. (2010). Virulence profile comparison between LEE-negative Shiga toxin-producing *Escherichia coli* (STEC) strains isolated from cattle and humans. *Vet. Microbiol.* 143, 307–313. doi: 10.1016/j.vetmic.2009.11.028
- Grisaru, S. (2014). Management of hemolytic-uremic syndrome in children. *Int. J. Nephrol. Renovasc. Dis.* 7, 231–239. doi: 10.2147/IJNRD.S41837
- Guinée, P., Jansen, W., Wadström, T., and Sellwood, R. (1981). “*Escherichia coli* associated with neonatal diarrhoea in piglets and calves,” in *Laboratory Diagnosis in Neonatal Calf and Pig Diarrhoea*, Vol. 13, eds P. W. de Leeuw and P. A. M. Guinée (Dordrecht: Springer), 126–162.
- Hauser, E., Mellmann, A., Semmler, T., Stoeber, H., Wieler, L. H., Karch, H., et al. (2013). Phylogenetic and molecular analysis of food-borne shiga toxin-producing *Escherichia coli*. *Appl. Environ. Microbiol.* 79, 2731–2740. doi: 10.1128/AEM.03552-12
- Karmali, M. A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., et al. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *J. Clin. Microbiol.* 41, 4930–4940. doi: 10.1128/JCM.41.11.4930-4940.2003
- Kaufmann, M., Zweifel, C., Blanco, M., Blanco, J., Blanco, J., Beutin, L., et al. (2006). *Escherichia coli* O157 and non-O157 Shiga toxin-producing *Escherichia coli* in fecal samples of finished pigs at slaughter in Switzerland. *J. Food Prot.* 69, 260–266. doi: 10.4315/0362-028X.JFP-10-579
- Koláčková, I., Házová, K., Škočková, A., and Karpíšková, R. (2014). [Occurrence of Shiga toxigenic *Escherichia coli* strains in pigs and cattle at slaughterhouses in the Czech Republic in 2013.]. *Klin. Mikrobiol. Infekc. Lek.* 20, 36–39.
- Koohmaraie, M., Arthur, T., Bosilevac, J., Guerini, M., Shackelford, S., and Wheeler, T. (2005). Post-harvest interventions to reduce/eliminate pathogens in beef. *Meat Sci.* 71, 79–91. doi: 10.1016/j.meatsci.2005.03.012
- Lee, G. Y., Jang, H. I., Hwang, I. G., and Rhee, M. S. (2009). Prevalence and classification of pathogenic *Escherichia coli* isolated from fresh beef, poultry, and pork in Korea. *Int. J. Food Microbiol.* 134, 196–200. doi: 10.1016/j.ijfoodmicro.2009.06.013
- Leung, P., Yam, W., Ng, W., and Peiris, J. (2001). The prevalence and characterization of verotoxin-producing *Escherichia coli* isolated from cattle and pigs in an abattoir in Hong Kong. *Epidemiol. Infect.* 126, 173–179. doi: 10.1017/S0950268801005210
- Magwedere, K., Dang, H. A., Mills, E. W., Cutter, C. N., Roberts, E. L., and Debroy, C. (2013). Incidence of Shiga toxin-producing *Escherichia coli* strains in beef, pork, chicken, deer, boar, bison, and rabbit retail meat. *J. Vet. Diagn. Invest.* 25, 254–258. doi: 10.1177/1040638713477407
- Martin, A., and Beutin, L. (2011). Characteristics of Shiga toxin-producing *Escherichia coli* from meat and milk products of different origins and association with food producing animals as main contamination sources. *Int. J. Food Microbiol.* 146, 99–104. doi: 10.1016/j.ijfoodmicro.2011.01.041
- Meng, Q., Bai, X., Zhao, A., Lan, R., Du, H., Wang, T., et al. (2014). Characterization of Shiga toxin-producing *Escherichia coli* isolated from healthy pigs in China. *BMC Microbiol.* 14:5. doi: 10.1186/1471-2180-14-5
- Mohlathole, R. P., Madoroba, E., Muchadeyi, F. C., Chimonyo, M., Kanengoni, A. T., and Dzomba, E. F. (2013). Virulence profiles of enterotoxigenic, shiga toxin and enteroaggregative *Escherichia coli* in South African pigs. *Trop. Anim. Health Prod.* 45, 1399–1405. doi: 10.1007/s11250-013-0377-4
- Niewerth, U., Frey, A., Voss, T., Le Bouguénec, C., Baljer, G., Franke, S., et al. (2001). The Aida autotransporter system is associated with edema disease and postweaning diarrhea. *Clin. Diagn. Lab. Immunol.* 8, 143–149.
- Padola, N. L., Sanz, M. E., Blanco, J. E., Blanco, M., Blanco, J., Etcheverría, A. A. I., et al. (2004). Serotypes and virulence genes of bovine Shiga-toxigenic *Escherichia coli* (STEC) isolated from a feedlot in Argentina. *Vet. Microbiol.* 100, 3–9. doi: 10.1016/S0378-1135(03)00127-5
- Parma, A., Sanz, M., Blanco, J., Blanco, J., Viñas, M., Blanco, M., et al. (2000). Virulence genotypes and serotypes of verotoxigenic *Escherichia coli* isolated from cattle and foods in Argentina. *Eur. J. Epidemiol.* 16, 757–762. doi: 10.1023/A:1026746016896
- Paton, A. W., and Paton, J. C. (2002). Direct detection and characterization of Shiga toxigenic *Escherichia coli* by multiplex PCR for stx1, stx2, eae, ehxA, and saa. *J. Clin. Microbiol.* 40, 271–274. doi: 10.1128/JCM.40.1.271-274.2002
- Paton, A. W., Srimanote, P., Woodrow, M. C., and Paton, J. C. (2001). Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. *Infect. Immun.* 69, 6999–7009. doi: 10.1128/IAI.69.11.6999-7009.2001
- Paton, J. C., and Paton, A. W. (1998). Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Clin. Microbiol. Rev.* 11, 450–479.
- Pradel, N., Livrelli, V., De Champs, C., Palcoux, J.-B., Reynaud, A., Scheut, F., et al. (2000). Prevalence and characterization of Shiga toxin-producing *Escherichia coli* isolated from cattle, food, and children during a one-year prospective study in France. *J. Clin. Microbiol.* 38, 1023–1031.
- Restieri, C., Garriss, G., Locas, M.-C., and Dozois, C. M. (2007). Autotransporter-encoding sequences are phylogenetically distributed among *Escherichia coli* clinical isolates and reference strains. *Appl. Environ. Microbiol.* 73, 1553–1562. doi: 10.1128/AEM.01542-06
- Rivas, M., Padola, N. L., Lucchesi, P. M., Masana, M., and Torres, A. (2010). “Diarrheagenic *Escherichia coli* in Argentina,” in *Pathogenic Escherichia coli in Latin America*, ed. A. G. Torres (Oak Park, IL: Bentham Science Publishers), 142–161.
- Roberts, T., Hudson, W., Whelehan, O., Simonsen, B., Olgaard, K., Labots, H., et al. (1984). Number and distribution of bacteria on some beef carcasses at selected abattoirs in some member states of the European Communities. *Meat Sci.* 11, 191–205. doi: 10.1016/0309-1740(84)90037-8
- SENASA (2002). “Circular 3496/02,” in *Animal*. Argentina: Servicio Nacional de Sanidad y Calidad Agroalimentaria.
- Sonntag, A.-K., Bielaszewska, M., Mellmann, A., Dierksen, N., Schierack, P., Wieler, L. H., et al. (2005). Shiga toxin 2e-producing *Escherichia coli* isolates from humans and pigs differ in their virulence profiles and interactions with intestinal epithelial cells. *Appl. Environ. Microbiol.* 71, 8855–8863. doi: 10.1128/AEM.71.12.8855-8863.2005
- Szalo, I. M., Goffaux, F., Pirson, V., Piérard, D., Ball, H., and Mainil, J. (2002). Presence in bovine enteropathogenic (EPEC) and enterohaemorrhagic (EHEC)

- Escherichia coli* of genes encoding for putative adhesins of human EHEC strains. *Res. Microbiol.* 153, 653–658. doi: 10.1016/S0923-2508(02)01379-7
- Tarr, P. I., Bilge, S. S., Vary, J. C., Jelacic, S., Habeeb, R. L., Ward, T. R., et al. (2000). Iha: a novel *Escherichia coli* O157: H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. *Infect. Immun.* 68, 1400–1407. doi: 10.1128/IAI.68.3.1400-1407.2000
- Trotz-Williams, L. A., Mercer, N. J., Walters, J. M., Maki, A. M., and Johnson, R. P. (2012). Pork implicated in a Shiga toxin-producing *Escherichia coli* O157: H7 outbreak in Ontario, Canada. *Can. J. Public Health* 103, e322–e326.
- Tseng, M., Frataamico, P. M., Bagi, L., Delannoy, S., Fach, P., Manning, S. D., et al. (2014). Diverse virulence gene content of Shiga toxin-producing *Escherichia coli* from finishing swine. *Appl. Environ. Microbiol.* 80, 6395–6402. doi: 10.1128/AEM.01761-14
- Wells, T. J., Sherlock, O., Rivas, L., Mahajan, A., Beatson, S. A., Torpdahl, M., et al. (2008). EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157: H7 that contributes to adhesion and biofilm formation. *Environ. Microbiol.* 10, 589–604. doi: 10.1111/j.1462-2920.2007.01479.x
- Woodward, M., Carroll, P., and Wray, C. (1992). Detection of entero- and verocytotoxin genes in *Escherichia coli* from diarrhoeal disease in animals using the polymerase chain reaction. *Vet. Microbiol.* 31, 251–261. doi: 10.1016/0378-1135(92)90083-6
- Wu, Y., Hinenoya, A., Taguchi, T., Nagita, A., Shima, K., Tsukamoto, T., et al. (2010). Distribution of virulence genes related to adhesions and toxins in Shiga toxin-producing *Escherichia coli* strains isolated from healthy cattle and diarrheal patients in Japan. *J. Vet. Med. Sci.* 72, 589–597. doi: 10.1292/jvms.09-0557
- Yoon, S. C., Windham, W. R., Ladely, S., Heitschmidt, G. W., Lawrence, K. C., Park, B., et al. (2013). Differentiation of big-six non-O157 Shiga-toxin producing *Escherichia coli* (STEC) on spread plates of mixed cultures using hyperspectral imaging. *J. Food Meas. Charact.* 7, 47–59. doi: 10.4315/0362-028X.JFP-12-497
- Zhao, L., Chen, X., Xu, X., Song, G., and Liu, X. (2009). Analysis of the AIDA-I gene sequence and prevalence in *Escherichia coli* isolates from pigs with post-weaning diarrhoea and oedema disease. *Vet. J.* 180, 124–129. doi: 10.1016/j.tvjl.2007.10.021

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

Copyright © 2016 Colello, Cáceres, Ruiz, Sanz, Etcheverría and Padola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Inactivation of Uropathogenic *Escherichia coli* in Ground Chicken Meat Using High Pressure Processing and Gamma Radiation, and in Purge and Chicken Meat Surfaces by Ultraviolet Light

Christopher H. Sommers*, O. J. Scullen and Shiowshuh Sheen

Eastern Regional Research Center, United States Department of Agriculture, Agricultural Research Service, Wyndmoor, PA, USA

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Diego Garcia-Gonzalo,
Universidad de Zaragoza, Spain
Olin Silander,
Massey University, New Zealand

*Correspondence:

Christopher H. Sommers
christopher.sommers@ars.usda.gov

Specialty section:

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

Received: 25 November 2015

Accepted: 14 March 2016

Published: 14 April 2016

Citation:

Sommers CH, Scullen OJ
and Sheen S (2016) Inactivation
of Uropathogenic *Escherichia coli*
in Ground Chicken Meat Using High
Pressure Processing and Gamma
Radiation, and in Purge and Chicken
Meat Surfaces by Ultraviolet Light.
Front. Microbiol. 7:413.
doi: 10.3389/fmicb.2016.00413

Extraintestinal pathogenic *Escherichia coli*, including uropathogenic *E. coli* (UPEC), are common contaminants in poultry meat and may cause urinary tract infections after colonization of the gastrointestinal tract and transfer of contaminated feces to the urethra. Three non-thermal processing technologies used to improve the safety and shelf-life of both human and pet foods include high pressure processing (HPP), ionizing (gamma) radiation (GR), and ultraviolet light (UV-C). Multi-isolate cocktails of UPEC were inoculated into ground chicken which was then treated with HPP (4°C, 0–25 min) at 300, 400, or 500 MPa. HPP D₁₀, the processing conditions needed to inactivate 1 log of UPEC, was 30.6, 8.37, and 4.43 min at 300, 400, and 500 MPa, respectively. When the UPEC was inoculated into ground chicken and gamma irradiated (4 and –20°C) the GR D₁₀ were 0.28 and 0.36 kGy, respectively. The UV-C D₁₀ of UPEC in chicken suspended in exudate and placed on stainless steel and plastic food contact surfaces ranged from 11.4 to 12.9 mJ/cm². UV-C inactivated ca. 0.6 log of UPEC on chicken breast meat. These results indicate that existing non-thermal processing technologies such as HPP, GR, and UV-C can significantly reduce UPEC levels in poultry meat or exudate and provide safer poultry products for at-risk consumers.

Keywords: UPEC, high pressure processing, gamma radiation, ultraviolet light, chicken

INTRODUCTION

Escherichia coli are classified as commensal (natural microflora), or variants that cause disease such as intestinal pathogenic *E. coli* (iPEC) or extraintestinal (ExPEC) types. Groups of ExPEC include Neonatal Meningococcal *E. coli* (NMEC), Avian Pathogenic *E. coli*, (APEC), Sepsis-associated Pathogenic *E. coli* (SEPEC) and Uropathogenic *E. coli* (UPEC) (Mitchell et al., 2015). *E. coli* such as ExPEC (UPEC) are responsible for 75–95% of urinary tract infections (UTI) and uncomplicated cystitis and pyelonephritis (Nordstrom et al., 2013). Fifty percent of women will contract one UTI in their lifetime, and 25% will have a recurrent UTI (Minardi et al., 2011; Bao et al., 2014). The number of UTI in the US is ca. 6–8 million annually, with ca. 100, 000 hospitalizations, ca. 23,000 deaths, and a health care burden

of ca. \$3.5 billion (Nordstrom et al., 2013). The mechanism for contraction of a UTI is transfer of contaminated feces from the gastrointestinal tract to the urethra, and isolates associated with UTI invariably match the individual's fecal microflora (Moreno et al., 2008).

The idea that extraintestinal foodborne pathogens such as the ExPEC might be responsible for UTI in humans is relatively new, and it has long been suspected they may be associated with illness outbreaks (Markland et al., 2015). The presence of ExPEC in poultry meat has been firmly established (Johnson et al., 2005; Mitchell et al., 2015). Studies have compared ExPEC isolates from food animals, food, and those from women with UTI and the incidence of ExPEC in poultry meat and have demonstrated both genetic similarity and identity between ExPEC from animals and food with those from humans with UTI (Cortes et al., 2010; Jakobsen et al., 2010a,b, 2012; Vincent et al., 2010; Bergeron et al., 2012; Mora et al., 2013). More importantly ExPEC isolated from animals and food can cause UTI in mouse model systems (Jakobsen et al., 2012).

Three non-thermal intervention technologies of interest to the meat and poultry processing industry, which are used commercially to improve food safety and extend shelf life, include high pressure processing (HPP), ionizing (gamma) radiation (GR) and ultraviolet light (UV-C) (Salvage, 2014). HPP subjects food to an elevated pressure of 100–1000 MPa typically at temperatures below 60°C. The mechanism by which HPP inactivates foodborne pathogens includes cell membrane and structure damage, ribosome dissociation, dissociation of DNA, and enzyme inactivation (Campus, 2010; Simonin et al., 2012). GR inactivates microorganisms by damaging their DNA indirectly through radiolysis of water and induction of oxidative damage or direct damage through breakage of the phosphodiester backbone in addition to oxidative damage to proteins and cell membranes (Taub et al., 1979; Diehl, 1995). UV-C kills microorganisms through induction of cyclobutane pyrimidine dimers and 6-4 photoproducts in addition to protein damage (Krisko and Radman, 2010; Rastogi et al., 2010).

The purpose of this study was to determine the HPP and GR inactivation kinetics for ExPEC (UPEC) inoculated in ground chicken as well as the UV-C inactivation kinetics on poultry meat surfaces and in chicken purge on food contact surfaces. To the authors knowledge this is the first study to examine the inactivation kinetics of ExPEC in a food system.

MATERIALS AND METHODS

Chicken

Ground chicken (92% lean) was freshly prepared and purchased at a local wholesaler (Lansdale, PA, USA) and evenly portioned into 90 g aliquots in polynylon pouches (Uline, Inc., Philadelphia, PA, USA), vacuum sealed to 50 millibars using a Multi-Vac A300 packager (Multi-Vac Inc., Kansas City, MO, USA) and then frozen (−70°C). The ground chicken was tested for presence of *E. coli* as described below and it was <1 CFU/g. Multiple chicken lots were tested and one with low *E. coli* levels was

selected. Boneless skinless chicken breast and chicken skin was obtained fresh from a local butcher. Chicken purge was obtained from a local poultry processor and frozen (−70°C) until ready for use.

E. coli Isolates

The *E. coli* isolates were obtained from the American Type Culture Collection (Manassas, VA, USA). These include 700414, 700415, 700416, 700417, 700336, and BAA-1161 (<http://www.atcc.org>), which were isolated from women with UTI. Multi-isolate cocktails of the pathogens were used as recommended for appropriate validation of non-thermal processing technologies (National Advisory Committee on Microbiological Criteria for Food [Nacmcf], 2006). The individual isolates were prescreened for resistance to HPP, GR and UV prior to use, and the D₁₀ were consistent with results for our previous studies with iPEC (Sheen et al., 2015; Sommers et al., 2015; Sommers et al., unpublished data).

E. coli Growth and Inoculation

The *E. coli* were cultured independently in 20 ml Tryptic Soy Broth (TSB) without dextrose to avoid development of acid resistance (BD-Difco, Sparks, MD, USA) using 50 ml sterile tubes at 37°C (150 rpm) for 18–24 h using a New Brunswick Model G34 Environmental Shaker (New Brunswick, Edison, NJ, USA). The bacteria were then sedimented by centrifugation (1,200 × g, Hermle Model Z206A, Hermle Labortechnik, Germany) and resuspended as a cocktail in 20 ml sterile 0.1% peptone water (SPW, BD-Difco).

Thawed ground chicken (10 g) was aliquoted into 2 oz. Nasco (Ft. Atkinson, WI, USA) Whirl-Pak bags, inoculated with 0.1 ml of UPEC, mixed manually for 1 min, and then sealed using the Multi-Vac A300 Packager. The final concentration of UPEC in the ground chicken was ca. 8–9 log CFU/g. The sample bags were then sealed in a second bag and stored at 4°C until HPP treatment or gamma radiation (ca. 2 h).

High Pressure Processing Treatment

High pressure processing was performed using a laboratory scale pressure unit (Mini Food lab FPG5620, Stansted Fluid Power Ltd., Essex, UK), comprised of a double-jacketed thick-wall stainless steel cylinder (approximate volume of 0.3 L) having an internal stainless steel sample holder of 25.4 mm × 254 mm (diameter × length). The thick-wall cylinder was maintained at a set-point temperature in which heat transfer fluid continuously circulated from a refrigerated liquid chiller (Proline RP 855, Lauda, Germany). The pressure come-up rate was 100 MPa per 15 s (or 6.7 MPa/s) and the release rate was 100 MPa per 9 s (or 11.1 MPa/s). Samples were pressure-treated at 500, 400, and 300 MPa (4°C) at 5 min intervals for up to 25 min. The initial temperature in the processing chamber was ca. 4°C and did not exceed a maximum of 35°C during the HPP treatment. Keeping the chamber temperature low (ca. 4°C) prevents compression heating induced thermal effects from interfering with HPP inactivation kinetic determination (Sheen et al., 2015). The chamber temperature was monitored by the built-in sensor (a

T-type thermal couple device). The thermal sensor was immersed in the working chamber near food samples filled with the recirculation fluid.

Gamma Radiation

A Lockheed Georgia Company (Marietta, GA, USA) self-contained ^{137}Cs irradiator, with a dose rate of 0.065 kGy/min, was used for all exposures. The radiation source consisted of 23 individually sealed source pencils in an annular array. The 22.9 cm \times 63.5 cm cylindrical sample chamber was located central to the array when placed in the operating position. Inoculated samples were placed vertically and centrally in the sample chamber, using a 4 mm thick polypropylene bucket, to ensure a good dose uniformity ($\text{DUR} < 1.1:1.0$). The temperature during irradiation (4°C) was monitored by thermocouple and maintained (4 or -20°C) by introduction of the gas phase from a liquid nitrogen source directly into the top of the sample chamber. The radiation doses were at 0.3 and 0.6 kGy increments at 4 or -20°C , respectively. The absorbed dose was verified using temperature tempered 5 mm alanine pellets that were then measured using a Bruker eScan EPR Analyzer (Bruker, Billerica, MA, USA).

Exposure to Ultraviolet Light

A custom built UV-C apparatus ($2 \text{ mW}/\text{cm}^2$) (Sommers et al., 2010) was used to treat chicken purge inoculated with UPEC on stainless steel (304 L), High Density Polypropylene (HDPP) and High Density Polyethylene (HDPE) coupons ($5 \times 10 \text{ cm}$), and the foods themselves. Chicken purge was thawed in a refrigerator overnight and 0.5 ml of UPEC cocktail inoculated into 4.5 ml chicken exudates was then mixed by vortexing for 30 s. One hundred microliter of inoculated purge was placed on the coupons which were then spread to a $4 \text{ cm} \times 4 \text{ cm}$ area using an inoculating loop. The coupons were placed in a refrigerator for 30 min and then placed on a cold pack (4°C) for UV-C exposure. The UV-C intensity exposure times were 0, 10, 20, 30, 40, 50, and 60 s for UV-C doses of 20, 40, 60, 80, 100, and $120 \text{ mJ}/\text{cm}^2$.

For chicken meat and skin $4 \times 4 \text{ cm}$ sections (ca. 1 mm thick) of boneless skinless chicken breast were placed in sterile petri dishes and inoculated with 0.1 ml of chicken purge which was then spread onto the surface ($4 \text{ cm} \times 4 \text{ cm}$) using an inoculating loop, and then incubated for 30 min in a refrigerator (4°C) prior to treatment with UV-C. The samples were placed on cold packs prior to UV-C treatment. The UV-C intensity exposure times were 0, 10, 20, 30, 40, 50, and 60 s for UV-C doses of 20, 40, 60, 80, 100, and $120 \text{ mJ}/\text{cm}^2$.

UV-C intensity was monitored using a calibrated UVX Radiometer (UVP Inc., Upland, CA, USA). The temperature of the room was approximately 20°C during the exposure to UV-C, and the food temperature did not increase to more than 30°C at the end of the process as measured using an infrared thermometer.

Recovery of the Surviving *E. coli*

The individual ground chicken samples were added to 90 ml of 0.1% PW and then stomached for 2 min (Model Bag Mixer 100W,

Inter science Co., France). The coupons with 0.1 ml exudate were placed in stomacher bags with 9.9 ml SPW and hand massaged for 1 min. For recovery of UPEC 1.0 mL, after proper decimal dilutions, was placed on duplicate *E. coli*/coliform PetrifilmTM (3M Microbiology Products Co., St. Paul, MN, USA). The films were maintained at room temperature for 4 h to allow the injured cells to recover (Hsu et al., 2014) and then incubated at 37°C for 24 h. Colonies (CFU) were enumerated for determination of log reduction and D_{10} . Incubation for longer periods did not result in changes to the colony counts, an indicator of injured cell recovery.

Statistical Analysis

The mean plate counts of the treated samples (N) were divided by the average control plate counts (N_0) to give a survivor ratio (N/N_0). The $\log_{10} (N/N_0)$ of the ratios was then used for determination of D_{10} -values and other statistical analyses. D_{10} -values were determined by the reciprocal of the slope following linear regression as determined by least squares analysis (Diehl, 1995). Each experiment (D_{10} determination) was conducted independently three times. A minimum of five time points were used for determination of D_{10} . Statistical analysis functions of MS Excel (Microsoft Corp., Redmond, WA, USA) were used for routine calculations (D_{10} determination), descriptive statistics, analysis of variance (ANOVA, 95% confidence).

RESULTS AND DISCUSSION

High Pressure Processing

The HPP inactivation kinetics for the UPEC multi-isolate cocktail is shown in **Table 1** and **Figure 1**. As we have found previously for STEC the inactivation kinetics was first order in nature. The HPP D_{10} of the UPEC in refrigerated (4°C) ground chicken was ca. 30.6, 8.37, and 4.43 min at 300, 400, and 500 MPa, respectively. HPP treatment using 300 MPa was ineffective as a treatment as ca. 1 log was inactivated at that pressure. When we compare the results of this study with those from previous HPP studies for

TABLE 1 | D_{10} values for uropathogenic *Escherichia coli* in ground chicken and chicken purge.

Technology	Parameter	D_{10} (SEM)
High pressure processing	300 MPa	30.6 (± 0.12) min
	400 MPa	8.37 (± 1.06) min
	500 MPa	4.4 (± 0.12) min
Gamma radiation	4°C	0.28 (± 0.01) kGy
	-20°C	0.36 (± 0.01) kGy
Ultraviolet light (chicken purge)	Stainless steel	11.9 (± 0.49) mJ/cm^2
	HDPP	11.4 (± 0.47) mJ/cm^2
	HDPE	12.9 (± 0.59) mJ/cm^2

D_{10} for HPP, GR, and ultraviolet light are shown with the standard error of the mean in parenthesis. Each experiment was conducted independently three times ($n = 3$). Each HPP D_{10} was significantly different than the others, as were the GR D_{10} (ANOVA, $p < 0.05$). There was no difference ($p > 0.05$) between the UV-C D_{10} for food contact surfaces.

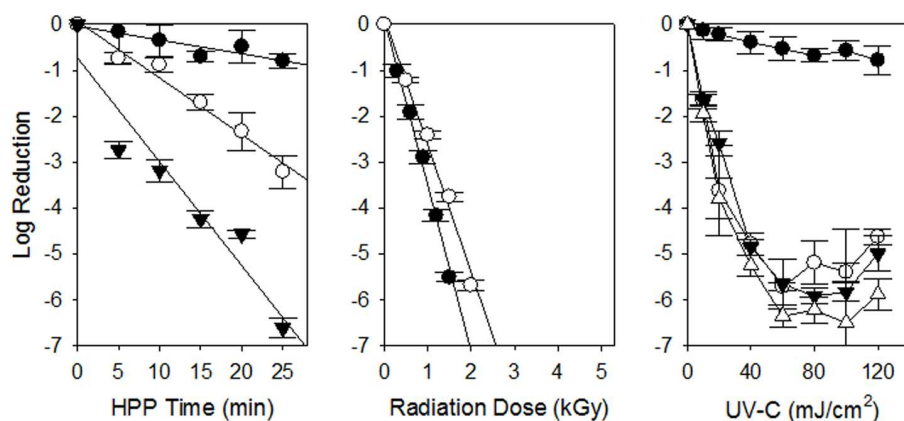


FIGURE 1 | Inactivation of uropathogenic *Escherichia coli* on chicken meat and chicken purge by non-thermal processing technologies. HPP 300 (●), 400 (○) and 500 (▼) MPa are shown as well as gamma radiation at 4 (●) and -20 (○) °C. For UV-C inactivation of UPEC on chicken breast meat (●), and chicken exudates on SS (○), HDPE (▼) and HDPP (△) are shown. Each experiment was conducted independently three times ($n = 3$). The standard error of the mean is shown as error bars. The linear regressions are shown as solid lines.

inactivation of STEC the results are similar. Sheen et al. (2015) found the mean HPP D_{10} (350 MPa) of 39 STEC isolates from illness outbreaks to be ca. 9.25 min while those from animals and environmental sources was ca. 10.4 min when suspended in 80% lean ground beef (350 MPa, 4°C). Hsu et al. (2014) found that 450 MPa (15 min, 4°C) inactivated 5.5–6.9 log of STEC in 77% lean ground beef while 350 MPa inactivated ca. 3.2–4.7 log. Jiang et al. (2015) was able to inactivate 3–4 log of STEC with HPP using multiple 1 min cycles at 400 MPa. Our results obtained using the UPEC were similar to those we have and others have obtained in the STEC suspended in ground beef.

Gamma Radiation

When the UPEC cocktail was suspended in ground chicken and treated with gamma radiation the GR D_{10} was ca 0.28 kGy at refrigeration (4°C) temperature (Figure 1, Table 1). These results are similar to those obtained by Sommers et al. (2015) which found the GR D_{10} of STEC associated with illness outbreaks to be ca. 0.27 kGy when suspended in refrigerated 80% lean ground beef. Sommers and Fan (2012) reviewed the studies for inactivation of *E. coli* O157:H7 in refrigerated ground beef in which the GR D_{10} ranged from 0.013 to 0.37 kGy. GR D_{10} for microorganisms irradiated in frozen foods are typically higher than that in refrigerated foods due to the limitation of indirect DNA damage through immobility of hydroxyl radicals produced by the radiolysis of water in the frozen state (Bruns and Maxcy, 1979; Taub et al., 1979). Lopez-Gonzalez et al. (1999) found the D_{10} for *E. coli* O157:H7 suspended in frozen beef (-15°C) beef to be 0.62 kGy. Thayer and Boyd (2001) found the GR D_{10} of *E. coli* O157:H7 in frozen ground beef (-20°C) to be 0.98 kGy. Black and Jaczynski (2006) obtained D_{10} of 0.33 and 0.35 kGy for *E. coli* O157:H7 in frozen (-20°C) ground beef and chicken, respectively. It appears the radiation doses needed to inactivate STEC in refrigerated and frozen meat and poultry products should also control the UPEC.

Ultraviolet Light

In this study our objective was to calculate a UV-C D_{10} value for the UPEC suspended in chicken exudate on SS, HDPP, and HDPE surfaces. The UV-C D_{10} for UPEC is shown in Table 1 and Figure 1. The D_{10} was calculated from the linear portion of the survival curve (0–60 mJ/cm²) and ranged from 11.4 to 12.9 mJ, cm² ($p > 0.05$, ANOVA). As with previous studies complete inactivation of microorganisms with UV-C is difficult because of shadowing by particulates in purge. The D_{10} for UPEC in purge obtained were very similar to those obtained with STEC suspended in veal purge (Sommers et al., unpublished data), as well as other foodborne pathogens (Sommers et al., 2012; Sommers and Sheen, 2015). A relatively low UV-C dose of 100 mJ/cm² should be able to inactivate ≥ 5 log of UPEC in chicken purge on food contact surfaces.

When we inoculated the UPEC onto skinless chicken meat we obtained ca. 0.6 (± 0.19), respectively, which was significantly reduced from the untreated controls ($p < 0.05$) which is consistent with previous results from our group as well as other researchers (Stermer et al., 1987; Sumner et al., 1996; Sommers et al., 2010). When the UPEC were inoculated onto chicken skin we did not obtain a significant reduction, which is again consistent with results we have obtained using other foodborne pathogens on chicken skin (Stermer et al., 1987; Sumner et al., 1996; Sommers et al., 2010). The reduced inactivation of the UPEC on skin and meat surfaces is expected due to the surface topology and shielding of the UPEC from UV-C (Gardner and Shama, 2000).

CONCLUSION

Our results indicate the HPP, GR, and UV-C inactivation kinetics of the UPEC are similar to our historical results for the STEC in meat and meat purge. The processing conditions used to control STEC should have similar effects on the UPEC.

DISCLAIMER

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

AUTHOR CONTRIBUTIONS

SS contributed to collection of high pressure processing data. OS was responsible for collection of UV-C data, CS was study director and designed the study, was responsible for data

collection and analysis, and was responsible for manuscript completion.

FUNDING

This Project was funded by USDA-ARS National Program 108 Food Safety Project No. 8072-42000-073-00D.

ACKNOWLEDGMENT

We would like to thank Jennifer Cassidy for editing this manuscript.

REFERENCES

- Bao, Y., Welk, B., Reid, G., and Burton, J. (2014). "Chapter 14-Role of the microbiome in the recurrent urinary tract infection," in *Novel Insights into Urinary Tract Infections and their Management*, ed. T. Matsumoto (London: Future Medicine Ltd.).
- Bergeron, C., Prussing, C., Boerlin, P., Daignault, L., Reid-Smith, R., Zhanel, G., et al. (2012). Chicken as a reservoir for extraintestinal pathogenic *Escherichia coli* in humans. *Canada. Emerg. Infect. Dis.* 18, 415–421. doi: 10.3201/eid1803.111099
- Black, J., and Jaczynski, J. (2006). Temperature effect on inactivation kinetics of *Escherichia coli* O157:H7 by electron beam in ground beef, chicken breast meat, and trout fillets. *J. Food Sci.* 71, M221–M227. doi: 10.1111/j.1750-3841.2006.00105.x
- Bruns, M., and Maxcy, R. (1979). Effect of irradiation temperature and drying on survival of highly radiation resistant bacteria in complex menstrua. *J. Food Sci.* 44, 1743–1746.
- Campus, M. (2010). High pressure processing of meat, meat products and seafood. *Food Eng. Rev.* 2, 256–273. doi: 10.1007/s12393-010-9028-y
- Cortes, P., Blane, V., Mora, A., Dahbi, G., Blanco, J., Blanco, M., et al. (2010). Isolation and characterization of potentially pathogenic antimicrobial resistant *Escherichia coli* from chicken and pig farms in Spain. *Appl. Environ. Microbiol.* 76, 2799–2805. doi: 10.1128/AEM.02421-09
- Diehl, J. (1995). *Safety of Irradiated Food*, 2nd Edn. (New York, NY: Marcel Dekker), 93–98.
- Gardner, D., and Shama, G. (2000). Modeling UV-induced inactivation of microorganisms on surfaces. *J. Food Prot.* 63, 63–70.
- Hsu, H.-Y., Sheen, S., Sites, J., Huang, L., and Wu, J. (2014). Effect of high pressure treatment on the survival of Shiga toxin-producing *Escherichia coli* in strawberry puree. *Food Microbiol.* 40, 25–30. doi: 10.1016/j.fm.2013.11.019
- Jakobsen, L., Garneau, P., Bruant, G., Harel, J., Olsen, S., Porsbo, L., et al. (2012). Is *Escherichia coli* urinary tract infection a zoonosis? Proof of direct link with production animals and meat? *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 1121–1129. doi: 10.1007/s10096-011-1417-5
- Jakobsen, L., Kurbasic, A., Skjøl-Rasmussen, L., Ejrnæs, K., Porsbo, L., Pedersen, K., et al. (2010a). *Escherichia coli* isolates from broiler chicken meat broiler chickens, pork, and pigs share phylogroups and antimicrobial resistance with community-dwelling humans and patients with urinary tract infection. *Foodborne Path. Dis.* 7, 537–547. doi: 10.1089/fpd.2009.0409
- Jakobsen, L., Spangholm, D., Pedersen, K., Jensen, L., Emborg, H.-D., Agersø, Y., et al. (2010b). Broiler chickens, broiler chicken meat, pigs, and pork as a source of ExPEC related virulence genes and resistance in *Escherichia coli* isolates from community dwelling humans and UTI patients. *J. Food Microbiol.* 142, 264–272. doi: 10.1016/j.jfoodmicro.2010.06.025
- Jiang, Y., Scheinberg, A., Senevirathne, R., and Cutter, C. (2015). The efficacy of short and repeated high-pressure processing treatments on the reduction of non-O157:H7 Shiga-toxin producing *Escherichia coli* in ground beef patties. *Meat. Sci.* 102, 22–26. doi: 10.1016/j.meatsci.2014.12.001
- Johnson, J., Kuskowski, M., Smith, K., O'Bryan, T., and Tatini, S. (2005). Antimicrobial-resistant and extraintestinal pathogenic *Escherichia coli* in retail foods. 2005. *J. Infect. Dis.* 191, 1040–1049. doi: 10.1086/428451
- Krisko, A., and Radman, M. (2010). Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14373–14377. doi: 10.1073/pnas.1009312107
- Lopez-Gonzalez, V., Murano, P., Brennan, R., and Murano, E. (1999). Influence of various commercial packaging conditions on survival of *Escherichia coli* O157:H7 to irradiation by electron beam versus gamma rays. *J. Food Prot.* 62, 10–15.
- Markland, S., LeStange, K., Sharma, M., and Kniel, K. (2015). Old friends in new places: Exploring the role of extraintestinal *E. coli* in intestinal disease and foodborne illness. *Zoonoses Public Health* 62, 491–496. doi: 10.1111/zph.12194
- Minardi, D., d'Anzeo, G., Cantoro, D., Conti, A., and Muzzonigro, G. (2011). Urinary tract infections in women: etiology and treatment options. *Int. J. Gen. Med.* 4, 333–343. doi: 10.2147/IJGM.S11767
- Mitchell, M., Johnson, J., Johnston, B., Curtiss, R., and Mellata, M. (2015). Zoonotic potential of *Escherichia coli* isolates from retail chicken meat products and eggs. *Appl. Environ. Microbiol.* 81, 1177–1187. doi: 10.1128/AEM.03524-14
- Mora, A., Viso, S., Lopez, C., Pilar Alonso, M., Garcia-Garrote, F., Dabhi, G., et al. (2013). Poultry as reservoir for extraintestinal pathogenic *Escherichia coli* O45:K1:H7-B2-ST95 in humans. *Vet. Microbiol.* 167, 506–512. doi: 10.1016/j.vetmic.2013.08.007
- Moreno, E., Andreu, A., Pigrau, C., Kuskowski, M., Johnson, J., and Prats, G. (2008). Relationship between *Escherichia coli* strains causing acute cystitis in women and the fecal *E.coli* population of the host. *J. Clin. Microbiol.* 46, 2529–2534. doi: 10.1128/JCM.00813-08
- National Advisory Committee on Microbiological Criteria for Food [Nacmcf]. (2006). Requisite scientific parameters for establishing the equivalence of alternative methods of pasteurization. *J. Food Prot.* 69, 1190–1216.
- Nordstom, L., Liu, P., and Price, L. (2013). Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. *Front. Microbiol.* 4:29. doi: 10.3389/fmicb.2013.00029
- Rastogi, R., Kumar, A., Tyagi, M., and Sinha, R. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J. Nucleic Acids* 2010, 592980. doi: 10.4061/2010/592980
- Salvage, B. (2014). *Strengthening Meat Safety. MeatPoultry.com: The Business Journal for Meat and Poultry Processors*. Available at: <http://www.meatpoultry.com/Writers/Bryan%20Salvage/Strengthening%20meat%20safety.aspx?cck=1>. [accessed 20 August 2015].
- Sheen, S., Cassidy, J., Scullen, O., and Sommers, C. (2015). Inactivation of a genetically diverse set of Shiga toxin-producing *Escherichia coli* in ground beef using high pressure processing. *Food Microbiol.* 52, 84–87. doi: 10.1016/j.fm.2015.07.001
- Simonin, H., Duranton, F., and de Lamballerie, M. (2012). New insights into the high-pressure processing of meat and meat products. *Comp. Rev. Food Sci. Food Safety* 11, 285–306. doi: 10.1111/j.1541-4337.2012.00184.x
- Sommers, C., and Fan, X. (2012). "Irradiation of ground beef and fresh produce," in *Nonthermal Processing Technologies for Food*, eds H. Q. Zhang, G. V.

- Barbosa-Cánovas, V. M. Bala Balasubramaniam, C. P. Dunne, D. F. Farkas, J. T. C. Yuan (Hoboken, NJ: Wiley-Blackwell), 236–246.
- Sommers, C., Rakjowski, K., Scullen, O., Cassidy, C., Fratamico, P., and Sheen, S. (2015). Inactivation of Shiga Toxin-Producing *Escherichia coli* in lean ground beef by gamma irradiation. *Food Microbiol.* 49, 231–234. doi: 10.1016/j.fm.2015.02.013
- Sommers, C., Scullen, B., Paoli, G., and Bhaduri, S. (2012). Inactivation of *Francisella tularensis* Utah-112 on food and food contact surfaces by ultraviolet light. *J. Food Proc. Technol.* S11-002, 1–4. doi: 10.4172/2157-7110.S11-002
- Sommers, C., and Sheen, S. (2015). Inactivation of avirulent *Yersinia pestis* on food and food contact surfaces by ultraviolet light and freezing. *Food Microbiol.* 50, 1–4. doi: 10.1016/j.fm.2015.02.008
- Sommers, C., Sites, J., and Musgrove, M. (2010). Ultraviolet light (254 nm) inactivation of pathogens on foods and stainless steel surfaces. *J. Food Safety* 30, 70–79. doi: 10.1111/j.1745-4565.2010.00220.x
- Stermer, R., Lasater-Smith, M., and Brasington, C. (1987). Ultraviolet radiation-an effective bactericide for fresh meat. *J. Food Prot.* 50, 108–111.
- Sumner, S. S., Wallner-Pendleton, E., Froning, G., and Stetson, L. (1996). Inhibition of *Salmonella typhimurium* on agar medium and poultry skin by ultraviolet energy. *J. Food Prot.* 59, 319–321.
- Taub, I., Halliday, J., and Sevilla, M. (1979). Chemical reactions in proteins irradiated at subfreezing temperatures. *Ad. Chem. Serol.* 180, 109–140. doi: 10.1021/ba-1979-0180.ch006
- Thayer, D., and Boyd, G. (2001). Effect of irradiation temperature on inactivation of *Escherichia coli* O157:H7 and *Staphylococcus aureus*. *J. Food Prot.* 64, 1624–1626.
- Vincent, C., Boerlin, P., Daignault, D., Dozois, C., Dozois, C., Dutil, L., et al. (2010). Food reservoir for *Escherichia coli* causing urinary tract infections. *Emerg. Inf. Dis.* 16, 88–94. doi: 10.3201/eid1601.091118
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Sommers, Scullen and Sheen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Surveillance of Extended-Spectrum Beta-Lactamase-Producing *Escherichia coli* in Dairy Cattle Farms in the Nile Delta, Egypt

Sascha D. Braun^{1,2*}, Marwa F. E. Ahmed³, Hosny El-Adawy^{4,5}, Helmut Hotzel⁴, Ines Engelmänn^{1,2}, Daniel Weiß^{1,2}, Stefan Monecke^{1,2,6} and Ralf Ehrlich^{1,2}

¹ Alere Technologies GmbH, Jena, Germany, ² InfectoGnostics Research Campus, Jena, Germany, ³ Department of Animal Hygiene and Zoonoses, Faculty of Veterinary Medicine, Mansoura University, Mansoura, Egypt, ⁴ Institute of Bacterial Infections and Zoonoses, Friedrich-Loeffler-Institut, Jena, Germany, ⁵ Department of Poultry Disease, Faculty of Veterinary Medicine, Kafrelsheikh University, Kafr El-Sheikh, Egypt, ⁶ Institute for Medical Microbiology and Hygiene, Technical University of Dresden, Dresden, Germany

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Amy J. Mathers,
University of Virginia, USA
Brandon Luedtke,
United States Department of
Agriculture, USA

*Correspondence:

Sascha D. Braun
sascha.braun@clondia.com

Specialty section:

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

Received: 25 February 2016

Accepted: 15 June 2016

Published: 04 July 2016

Citation:

Braun SD, Ahmed MFE, El-Adawy H,
Hotzel H, Engelmänn I, Weiß D,
Monecke S and Ehrlich R (2016)
Surveillance of Extended-Spectrum
Beta-Lactamase-Producing
Escherichia coli in Dairy Cattle Farms
in the Nile Delta, Egypt.
Front. Microbiol. 7:1020.
doi: 10.3389/fmicb.2016.01020

Introduction: Industrial livestock farming is a possible source of multi-resistant Gram-negative bacteria, including producers of extended spectrum beta-lactamases (ESBLs) conferring resistance to 3rd generation cephalosporins. Limited information is currently available on the situation of ESBL producers in livestock farming outside of Western Europe. A surveillance study was conducted from January to May in 2014 in four dairy cattle farms in different areas of the Nile delta, Egypt.

Materials and Methods: In total, 266 samples were collected from 4 dairy farms including rectal swabs from clinically healthy cattle ($n = 210$), and environmental samples from the stalls ($n = 56$). After 24 h pre-enrichment in buffered peptone water, all samples were screened for 3rd generation cephalosporin-resistant *Escherichia coli* using Brilliance™ ESBL agar. Suspected colonies of putatively ESBL-producing *E. coli* were sub-cultured and subsequently genotypically and phenotypically characterized. Susceptibility testing using the VITEK-2 system was performed. All suspect isolates were genotypically analyzed using two DNA-microarray based assays: CarbDetect AS-1 and *E. coli* PanType AS-2 kit (ALERE). These tests allow detection of a multitude of genes and their alleles associated with resistance toward carbapenems, cephalosporins, and other frequently used antibiotics. Serotypes were determined using the *E. coli* SeroGenotyping AS-1 kit (ALERE).

Results: Out of 266 samples tested, 114 (42.8%) ESBL-producing *E. coli* were geno- and phenotypically identified. 113 of 114 phenotypically 3rd generation cephalosporin-resistant isolates harbored at least one of the ESBL resistance genes covered by the applied assays [*bla*CTX-M15 ($n = 105$), *bla*CTX-M9 ($n = 1$), *bla*TEM ($n = 90$), *bla*SHV ($n = 1$)]. Alarming, the carbapenemase genes *bla*OXA-48 ($n = 5$) and *bla*OXA-181 ($n = 1$) were found in isolates that also were phenotypically resistant to imipenem and meropenem. Using the array-based serogenotyping method, 66 of the 118 isolates (55%) could be genotypically assigned to O-types.

Conclusion: This study is considered to be a first report of the high prevalence of ESBL-producing *E. coli* in dairy farms in Egypt. ESBL-producing *E. coli* isolates with different underlying resistance mechanisms are common in investigated dairy cattle farms in Egypt. The global rise of ESBL- and carbapenemase-producing Gram-negative bacteria is a big concern, and demands intensified surveillance.

Keywords: ESBL, carbapenemases, *Escherichia coli*, Egypt, dairy cattle, microarray, genotype, CRE

INTRODUCTION

Extended-spectrum beta-lactamases (ESBLs) are mainly plasmid-encoded enzymes providing resistance to 3rd generation (3G) cephalosporins. These enzymes can be produced by a variety of different bacteria, such as *Enterobacteriaceae* or non-fermenting bacteria (Bradford, 2001; Giamarellou, 2005; Rawat and Nair, 2010; Shaikh et al., 2015). The most frequently found ESBL-producing species is *Escherichia coli* which often causes urinary tract infections, pneumonia or even sepsis in humans (Abraham et al., 2012). ESBL-producing *E. coli* has been broadly recognized in veterinary medicine as causative agents of mastitis in dairy cattle since the 2000s (Brinas et al., 2003; Haftu et al., 2012), but only a few studies exist that investigated the prevalence of ESBL-producing bacteria in livestock, showing their existence in sick and/or healthy cattle (Valentin et al., 2014; Dahms et al., 2015).

Unfortunately, there is no legislation in Egypt regulating the use of antibiotics (Dahshan et al., 2015). Antimicrobials such as tetracycline, quinolones, and beta lactams are still used in Egypt for growth promotion in animal feed and by veterinarians for the treatment and prevention of zoonotic diseases (WHO, 2013).

The CTX-M beta-lactamases, named for their greater activity against cefotaxime, are the most frequently detected ESBLs in livestock, and have been reported from different food-producing animals (Schmid et al., 2013; Brolund, 2014; Hansen et al., 2014). These animals also represent a source and/or a reservoir for ESBL-producing *E. coli* (Carattoli, 2008). Several studies indicate that these resistance genes are disseminated through the food chain or via direct contact between humans and animals (Schmid et al., 2013; Dahms et al., 2015). Data on ESBL-producing bacteria in food animals from Egypt are very limited. Therefore, the current study was conducted on four dairy cattle farms in different districts of Northern Egypt to assess the prevalence of ESBL-producing *E. coli* in dairy cattle and their environment.

MATERIALS AND METHODS

Farm Description and Sampling

In 2014, four dairy farms, three in Gamasa (GF1, GF2, GF5), and one in Damietta (D), were investigated (Figure 1). These farms were located in Nile Delta, Egypt in two different governorates (Damietta; Latitude N. 31°19', Longitude E. 31°81' and Dakahlia Latitude N. 31°25', Longitude E. 31°32'). The herd size ranged from 400, 600, 650, and 800 in GF1, GF2, GF5, and D, respectively. The cattle enrolled in this study were between 2 and 10 years old. Half of the farms housed dairy cattle in free-stall barns and half of them in tie-stall barns. In total, 266 samples

were collected from these farms. This included rectal swabs and milk samples from apparently healthy dairy cattle ($n = 210$), and environmental swab samples from water trough, feed and bedding ($n = 56$).

Bacterial Strains, Isolation, Identification, and Genomic DNA Extraction

All collected samples were enriched in buffered peptone water and cultivated on Brilliance™ ESBL agar (Oxoid, Wesel, Germany) for preliminary analysis for ESBL-producing *E. coli*. For further investigations, all suspected *E. coli* that grew on the selective medium were cultivated on tryptone yeast agar (Oxoid, Wesel, Germany). Presumptive characteristic *E. coli* isolates were identified by Gram staining and motility, and confirmed using a panel of biochemical tests (Triple Sugar Iron (TSI) agar, catalase, oxidase, H₂S production and sugar fermentation) and API 20 E systems (bioMérieux, France; ISO, 2001). All confirmed isolates were subsequently re-tested using an automated microdilution technique (VITEK-2, bioMérieux, Nürtingen, Germany) that covered the following antibiotics: imipenem, meropenem, cefotaxime, ceftazidime, cefuroxime-axetil, cefuroxime, piperacillin/tazobactam, ampicillin/sulbactam, ampicillin, gentamicin, tobramycin, ciprofloxacin, moxifloxacin, tetracycline, tigecycline, co-trimoxazol, and fosfomycin (VITEK-2 test card: AST-N289). Susceptibility tests for chloramphenicol, kanamycin, streptomycin, erythromycin and colistin were not performed in this study.

Genomic DNA from clonal isolates was extracted using the DNeasy Blood & Tissue kit (Qiagen GmbH, Hilden, Germany) according to manufacturer's instructions. When necessary, DNA was concentrated to at least 100 ng/μl using a SpeedVac centrifuge (Eppendorf, Hamburg, Germany) at room temperature with 1400 rpm and for 30 min. Five microliters of recovered genomic DNA were used directly for biotin-labeling and subsequent hybridization.

GenoSeroTyping and Antimicrobial Resistance Genotype

For all ESBL-producing *E. coli*, the serotype was determined using the *E. coli* SeroGenoTyping AS-1 kit. The antimicrobial resistance (AMR) genotype was detected by the CarbDetect AS-1 kit and all other resistance genes were detected by the *E. coli* PanType AS-2 kit (Alere Technologies GmbH, Jena, Germany). The data were automatically summarized by the "result collector," a software tool provided by Alere Technologies. An antibiotic resistance genotype was defined as a group of genes which have been described to confer resistance to a family of antibiotics (e.g.,

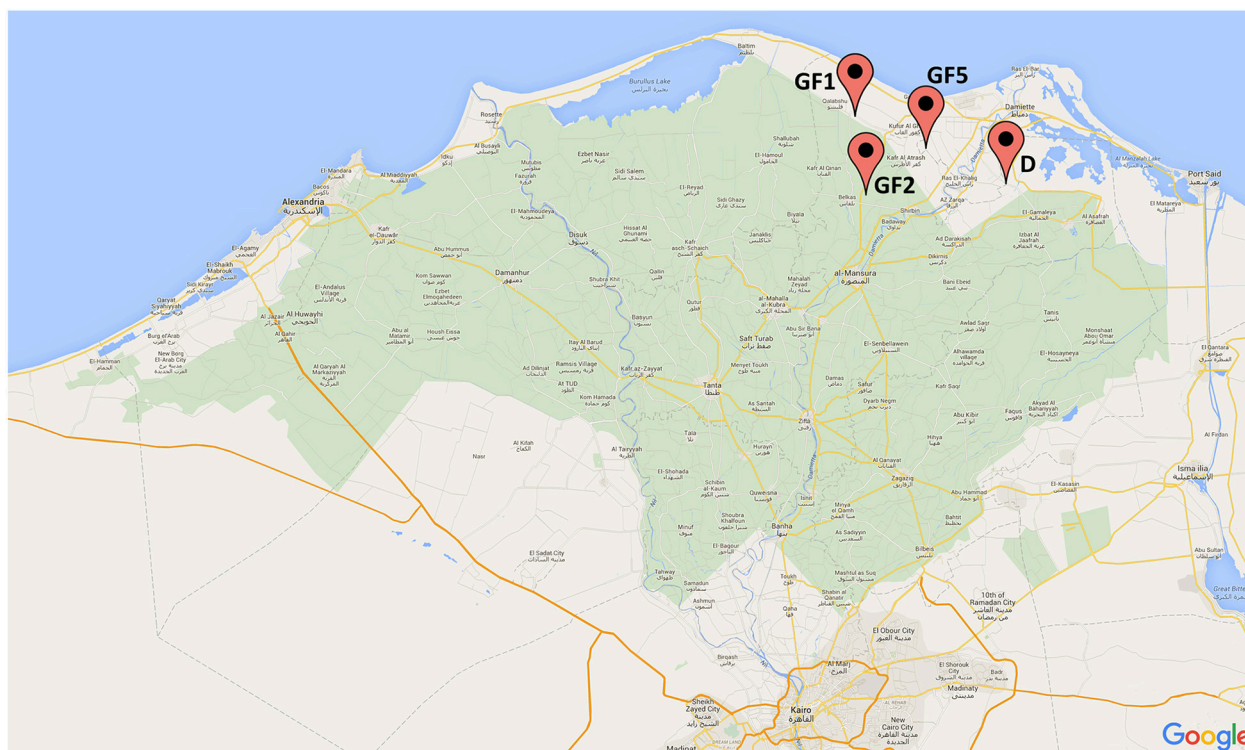


FIGURE 1 | Map of Nile Delta, Egypt with all locations of the dairy cattle farms from which the samples were collected (GF-Gamasa, D-Damietta).

the genotype “*bla*CTX-M1/15, *bla*TEM” confers resistance to 3G cephalosporins) (Table 2).

Multiplex Labeling, Hybridization, and Data Analysis

Extracted DNA was labeled by primer extension amplification using *E. coli* SeroGenoTyping AS-1, CarbDetect AS-1 or *E. coli* PanType AS-2 kits according to manufacturer's instructions. The procedure for multiplex labeling, hybridization and data analysis was described in detail by Braun et al. (2014). Briefly, internal labeling of the synthesized single stranded DNA resulted from the primer elongation of previously hybridized primers to the target genomic DNA, by using dUTP linked biotin as dideoxynucleotide triphosphate to be incorporated during synthesis. This procedure allowed site-specific internal labeling of the corresponding target region. The PCR protocol included 5 min of initial denaturation at 96°C, followed by 50 cycles with 20 s of annealing at 50°C, 40 s of elongation at 72°C, and 60 s of denaturation at 96°C (used device: Eppendorf Mastercycler gradient, Eppendorf, Hamburg, Germany). This reaction resulted in a multitude of specific linearly amplified, single-stranded, biotin-labeled DNA molecules for subsequent hybridization and detection using the DNA microarrays.

For hybridization procedures, the CarbDetect AS-1 and the *E. coli* PanType AS-2 kit were used according to manufacturer's instructions. CarbDetect ArrayStrips were placed in a thermomixer with an Alere ArrayStrip adapter (Quantifoil

Instruments, Jena, Germany) and subsequently washed with 200 µl of deionized water at 50°C with 550 rpm for 5 min and with 100 µl hybridization buffer C1 at 50°C with 550 rpm for 5 min. Liquids were always completely removed using a soft plastic pipette (e.g., BRANDT, #612-2856) to avoid any scratching of the chip surface. In a separate tube, 10 µl of previously labeled, single-stranded DNA was dissolved in 90 µl hybridization buffer C1. The hybridization was carried out at 50°C and 550 rpm for 1 h. After hybridization, the ArrayStrips were washed twice using 200 µl washing buffer C2 at 45°C for 10 min, shaking at 550 rpm. Peroxidase-streptavidin conjugate C3 was diluted 1:100 in buffer C4. A total of 100 µl of this mixture was added to each well of the ArrayStrip and subsequently incubated at 30°C and 550 rpm for 10 min. Thereafter, two washing steps with 200 µl C5 washing buffer were carried out at 550 rpm at 30°C for 5 min. The visualization was achieved by adding 100 µl of staining substrate D1 to the ArrayStrips, and signals were detected using the ArrayMate device (Alere Technologies GmbH). Finally, an automatically generated HTML-report was provided giving information on the presence or absence of antimicrobial resistance genes and the affiliation to one of the more common species.

Ethic Statement

An Ethic Statement is not necessary. The isolates were obtained by noninvasive rectal swabs and no animal experiments were carried out for this study.

RESULTS

Antimicrobial Resistance Genotype and Phenotype

Rectal swabs samples ($n = 210$) yielded 98 (46.6%) cultures and environmental samples ($n = 56$) yielded 16 (28.6%) cultures of putatively ESBL-producing *E. coli*. All 114 isolates were Gram-negative, motile, catalase positive, oxidase negative and indole-producing bacteria. Additionally, all isolates caused a decrease of pH and a color change of the TSI agar indicator and gas formation in the bottom of the test tube, and were therefore assigned as *E. coli*. In total, 113 (99.1%) phenotypically 3G cephalosporin-resistant isolates harbored at least one of the ESBL genes covered by the microarray (*bla*CTX-M15, *bla*CTX-M9, *bla*TEM, *bla*SHV; **Figure 2**). The carbapenemase gene *bla*OXA-48 was detected in five isolates (3.4%) and the carbapenemase gene *bla*OXA-181 (0.8%) was detected in one isolate (**Table 1**, **Figure 2**). These isolates showed a phenotypic resistance to imipenem and meropenem (**Table 2**). The total number of detected resistance genes is listed in **Table 1** and an overview to each isolate is given in **Figure 2**. The ESBL gene *bla*CTX-M1/15 was found in 103 isolates whereas *bla*CTX-M9 was only found in 9 isolates. Consensus sequences for *bla*TEM and *bla*SHV were found in 89 and 1 isolate, respectively. For all detected beta-lactamase genotypes, the phenotype was analyzed using the VITEK-2 instrument. The results are shown in **Table 2**. The concordance for the carbapenem resistance and ESBL genotype was 100%. In one phenotypic ESBL positive isolate, only the narrow spectrum beta-lactamase (NSBL) gene *bla*OXA-1 was found. Due to this unexpected phenotype the concordance for the NSBL genotype was only 60%.

Nine different aminoglycoside resistance genes were detected by the *E. coli* PanType AS-1 kit (**Table 1**). The combinations of these genes resulted in 14 different genotypes, whereas the most prevalent genotype was *aac*(6′)-*Ib* in combination with *aadA4* ($n = 25$). All isolates harboring this genotype were resistant to tobramycin, but three of them were susceptible to gentamicin (**Table 2**). Therefore, the concordance between genotype and phenotype was 94%. Five isolates, which harbored only *aac*(6′)-*Ib*, were resistant to all tested aminoglycosides. The second most frequent genotype was *aac*(3)-*IVa* ($n = 20$). All isolates harboring this gene were resistant to gentamicin and tobramycin and corresponded to 100% of the expected phenotype (**Table 2**). Six isolates harboring the gene *aphA* were susceptible to tobramycin and gentamicin. The detected phenotype corresponded 100% with the genotype, as the enzyme AphA does not mediate resistance against both aminoglycoside antibiotics tested (Ramirez and Tolmasky, 2010).

The most prevalent genotype for fluoroquinolone resistance was *qnrA1* followed by *qepA*. One isolate harboring *qnrA1* was sensitive to both quinolone antibiotics tested (97.0% concordance), but all isolates with detected *qepA* gene were resistant (100% concordance). Overall, 86 of 114 isolates were resistant to ciprofloxacin and 92 against moxifloxacin. Only in 56 ciprofloxacin resistant isolates a corresponding genotype was detected. Similar results were observed for the 92 moxifloxacin resistant isolates, where only in 68 isolates a corresponding

genotype was detected. The overall concordance of the detection of genes mediating fluoroquinolone resistance with phenotypic resistance was 79.0% (**Table 2**).

Resistance to co-trimoxazole is associated with *sul* and *dfrA* genes. All isolates with this gene combination were resistant to co-trimoxazole (**Table 2**). From 45 isolates without this gene combination, 34 were resistant. Therefore, the concordance of genotype and phenotype was 80.0%.

Of 114 isolates, one was resistant to fosfomycin. Such resistance is caused by mutations in ubiquitous genes (e.g., *murA* or *glpT*) or the loss of entire genes (e.g., *uhpA*) rather than by acquisition of distinct resistance markers and therefore the genotypes were not included into the test panel (Takahata et al., 2010; Li et al., 2015).

In summary, the overall concordance among all genotypes to expected phenotypes was 82.6% (**Table 2**).

Serotyping

For all 114 ESBL-producing *E. coli* isolates, the O- and H-types were identified using the *E. coli* SeroGenotyping AS-1 kit (**Figure 3**). For 63 (55.3%) isolates, genes encoding both O and H antigens were detected and for 51 isolates (44.7%) only the gene encoding the H antigen could be detected. The most prevalent serotype was O101:H10. This serotype was found in locations D, GF1, and GF2 and isolated mainly from rectal swabs, as one isolate belonging to this group was found in a water trough. The AMR genotype for O101:H10 isolates was rather uniform (**Figure 3**). Isolates of serotype O53:H18 with the carbapenemase gene *bla*OXA-48 were only found in farm GF1 and were isolated from rectal swabs. All isolates belonging to this group were identical with regard to their phenotype and genotypes (**Table 2**, **Figures 2, 3**). The serotype O8:H9 with *bla*OXA-181 was found only once in farm D from a rectal swab. For isolates where only the H-antigen H6 was found, a very similar AMR genotypes was detected. Such isolates were found in all investigated farms and sample types.

DISCUSSION

In 2014, four dairy farms in northern Egypt were investigated for ESBL-producing *E. coli*. In total, 210 clinically healthy dairy cattle were sampled using rectal swabs. Additionally, 56 environmental swabs were taken from different stall objects and screened for multi-drug resistant bacteria. All swabs were pre-cultured on Brilliance™ ESBL Agar, and in 114 of 266 samples (42.8%), ESBL-producing *E. coli* were detected. To analyze the underlying molecular AMR mechanism, all 114 isolates were genotyped using the multiplex microarray technique. The most frequently detected gene which mediated resistance against 3G cephalosporins was *bla*CTX-M1/15 (90.4%). The genes *bla*CTX-M9 (5.3%) and *bla*TEM (78%), which also mediate resistance to 3G cephalosporins, were also detected. However, both were usually found in combination with *bla*CTX-M1/15. *Bla*CTX-M9 was related to resistance to 3G cephalosporins in just one isolate as well as *bla*TEM in four isolates. A comparison to data similar to data from this study is difficult due to missing reports from Egypt. Recently, two reports from Germany are known describing the

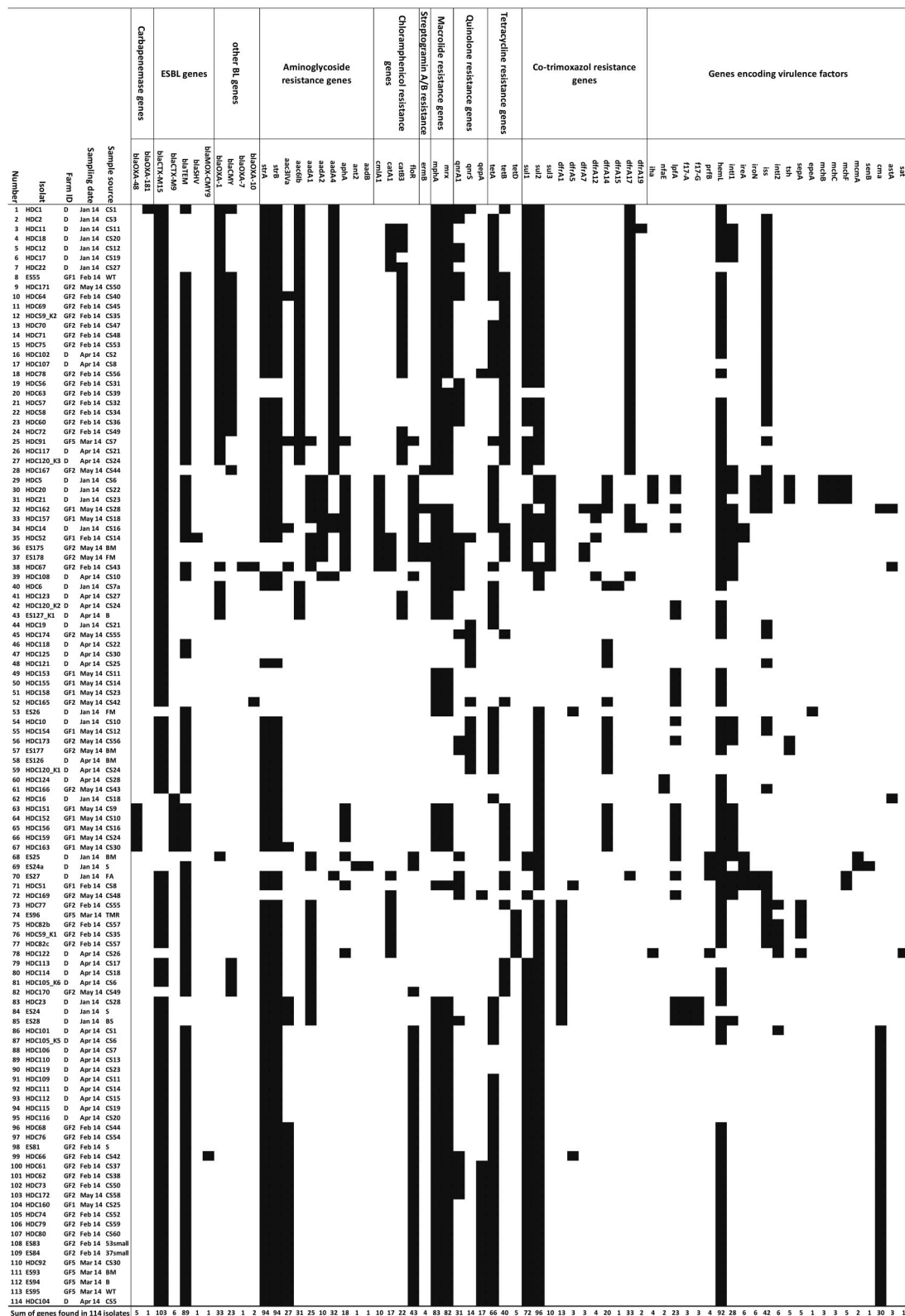


TABLE 1 | Antimicrobial resistance genes and their frequency in *E. coli* isolates detected by microarray.

Gene family	Genes	Accession No.	Description	Frequency
Carbapenemase	<i>bla</i> OXA-48	AY236073.2	Carbapenemase, class D carbapenem hydrolyzing beta-lactamase	5
	<i>bla</i> OXA-181	JN205800.1	Carbapenemase, class D carbapenem hydrolyzing beta-lactamase	1
ESBL (extended spectrum beta- lactamase)	<i>bla</i> CTX-M1/15	X92506.1	Class A extended-spectrum-beta-lactamase	103
	<i>bla</i> CTX-M9	AF174129.3	Class A extended-spectrum-beta-lactamase	6
	<i>bla</i> TEM	Consensus	Class A beta-lactamase	89
	<i>bla</i> SHV	Consensus	Class A beta-lactamase	1
	<i>bla</i> MOX-CMY9	AF381617.1	Class C beta-lactamase	1
NSBL (narrow spectrum beta-lactamase)	<i>bla</i> OXA-1	AY458016.1	Class D beta-lactamase	33
	<i>bla</i> OXA-7	AY866525.1	Class D beta-lactamase	1
	<i>bla</i> OXA-10	J03427.1	Class D beta-lactamase	2
	<i>bla</i> CMY	AB212086.1	Class C beta-lactamase	23
Aminoglycosides	<i>strA</i>	EF090911.1	Aminoglycoside-3''-phosphotransferase (locus A)	94
	<i>strB</i>	EF090911.1	Aminoglycoside-6''-phosphotransferase	94
	<i>aadA4</i>	Z50802.3	Aminoglycoside adenylyltransferase	32
	<i>aac</i> (6')-Ib	AM283490.1	Aminoglycoside 6'-N-acetyltransferase	31
	<i>aac</i> (3')-IVa	EU784152.1	Aminoglycoside 3'-N-acetyltransferase	27
	<i>aadA1</i>	EU704128.1	Aminoglycoside adenylyltransferase	25
	<i>aphA</i>	AY260546.3	Aminoglycoside 3'-phosphotransferase	18
	<i>aadA2</i>	EU704128.1	Aminoglycoside adenylyltransferase	10
	<i>aadB</i>	L06418.4	Aminoglycoside 2''-O-nucleotidyltransferase	1
Chloramphenicol	<i>floR</i>	AF252855.1	Florfenicol export protein	43
	<i>catB3</i>	AJ009818.1	Chloramphenicol acetyltransferase (group B)	22
	<i>catA1</i>	V00622.1	Chloramphenicol acetyltransferase (group A)	17
	<i>cmlA1</i>	EF113389.1	Chloramphenicol transporter	10
Streptogramin A/B	<i>ermB</i>	AB089505.1	rRNA adenine N-6-methyltransferase	4
Macrolides	<i>mphA</i>	AB038042.1	Macrolide 2'-phosphotransferase I	83
	<i>Mrx</i>	AB038042.1	Unknown (downstream to <i>mphA</i>)	82
Fluoroquinolones	<i>qnrA</i>	AY931018.1	Quinolone or fluoroquinolone resistance protein	31
	<i>qepA</i>	AM886293.1	Fluoroquinolone efflux pump	17
	<i>qnrS</i>	AM234722.1	Quinolone or fluoroquinolone resistance protein	14
Tetracycline	<i>tetA</i>	CP000971.1	Tetracycline resistance protein A, class A	66
	<i>tetB</i>	V00611.1	Tetracycline resistance protein A, class B	40
	<i>tetD</i>	X65876.1	Tetracycline resistance protein A, class D	5
Co-trimoxazole	<i>sul2</i>	DQ464881.1	Dihydropteroate synthetase type 2	96
	<i>sul1</i>	AJ698325.1	Dihydropteroate synthetase type 1	72
	<i>dfrA17</i>	AF169041.1	Dihydrofolate reductase type 17	33
	<i>dfrA14</i>	AJ313522.1	Dihydrofolate reductase type 14	20
	<i>dfrA1</i>	AJ884723.1	Dihydrofolate reductase type 1	13
	<i>sul3</i>	AJ459418.2	Dihydropteroate synthetase type 3	10
	<i>dfrA12</i>	AB154407.1	Dihydrofolate reductase type 12	4
	<i>dfrA5</i>	AB188269.1	Dihydrofolate reductase type 5	3
	<i>dfrA7</i>	AB161450.1	Dihydrofolate reductase type 7	3
	<i>dfrA19</i>	AJ310778.1	Dihydrofolate reductase type 19	2
	<i>dfrA15</i>	Z83311.1	Dihydrofolate reductase type 15	1

TABLE 2 | In detail comparison between the microarray-based genotype and the phenotype obtained by a VITEK-2 system.

AMR genotype detected by microarray ^a	Expected AMR phenotype	Frequency of genotype detected by microarray	Detected phenotype												Concordance of expected AMR phenotype (%)			
			Frequency of genotype detected by microarray						Detected phenotype									
			Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)				
<i>bla</i> OXA-48, <i>bla</i> CTX-M9, <i>bla</i> aTEM	Carbapenem resistant	5	MEM ^b 5	0	100	IMP 5	0	100	CTX 5	0	100	CAX 5	0	100	Amp 5	0	100	100
<i>bla</i> OXA-181, <i>bla</i> CTX-M1/15, <i>bla</i> aTEM, <i>bla</i> OXA-1, <i>bla</i> CMY	Carbapenem resistant	1	MEM 1	0	100	IMP 1	0	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
		6	6	0	100	6	0	100	6	0	100	6	0	100	6	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> aTEM	3G cephalosporin resistant ^c	53	MEM 0	53	100	IMP 0	53	100	CTX 53	0	100	CAX 53	0	100	Amp 53	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> aTEM, <i>bla</i> OXA-1, <i>bla</i> CMY	3G cephalosporin resistant	16	MEM 0	16	100	IMP 0	16	100	CTX 16	0	100	CAX 16	0	100	Amp 16	0	100	100
<i>bla</i> CTX-M1/15	3G cephalosporin resistant	12	MEM 0	12	100	IMP 0	12	100	CTX 12	0	100	CAX 12	0	100	Amp 12	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> OXA-1	3G cephalosporin resistant	10	MEM 0	10	100	IMP 0	10	100	CTX 10	0	100	CAX 10	0	100	Amp 10	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> aTEM, <i>bla</i> CMY	3G cephalosporin resistant	4	MEM 0	4	100	IMP 0	4	100	CTX 4	0	100	CAX 4	0	100	Amp 4	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> aTEM, <i>bla</i> OXA-1	3G cephalosporin resistant	3	MEM 0	3	100	IMP 0	3	100	CTX 3	0	100	CAX 3	0	100	Amp 3	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> CMY	3G cephalosporin resistant	1	MEM 0	1	100	IMP 0	1	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> MOX-CMY9, <i>bla</i> aTEM	3G cephalosporin resistant	1	MEM 0	1	100	IMP 0	1	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> aTEM, <i>bla</i> OXA-7	3G cephalosporin resistant	1	MEM 0	1	100	IMP 0	1	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
<i>bla</i> CTX-M1/15, <i>bla</i> aTEM, <i>bla</i> aSHV	3G cephalosporin resistant	1	MEM 0	1	100	IMP 0	1	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
<i>bla</i> CTX-M9	3G cephalosporin resistant	1	MEM 0	1	100	IMP 0	1	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
<i>bla</i> aTEM (consensus), <i>bla</i> CMY	3G cephalosporin resistant	1	MEM 0	1	100	IMP 0	1	100	CTX 1	0	100	CAX 1	0	100	Amp 1	0	100	100
<i>bla</i> aTEM (consensus)	3G cephalosporin resistant	3	MEM 0	3	100	IMP 0	3	100	CTX 3	0	100	CAX 3	0	100	Amp 3	0	100	100
		107	0	107	100	0	107	100	107	0	100	107	0	100	107	0	100	100

(Continued)

TABLE 2 | Continued

AMR genotype detected by microarray ^a	Expected AMR phenotype	Frequency of genotype detected by microarray	Detected phenotype												Concordance of expected AMR phenotype (%)								
			Frequency of genotype detected by microarray				Frequency of genotype detected by microarray				Frequency of genotype detected by microarray												
			Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)									
blaOXA-1	NSBL	1	MEM	0	1	100	IMP	0	1	100	CTX	1	0	0	CAZ	1	0	0	Amp	1	0	100	60
		1		0	1	100		0	1	100		1	0	0		1	0	0		1	0	100	60
aac3Va	Aminoglycoside resistant	20	CN	20	0	100	TOB	20	0	100													100
aac6Ib	Aminoglycoside resistant	5	CN	5	0	100	TOB	5	0	100													100
aadA1	Aminoglycoside resistant	11	CN	1	10	90	TOB	1	10	90													90
aadA1, aac3Va	Aminoglycoside resistant	1	CN	1	0	100	TOB	1	0	100													100
aadA2, aadA4	Aminoglycoside resistant	1	CN	0	1	0	TOB	0	1	0													0
aadA4	Aminoglycoside resistant	2	CN	1	1	50	TOB	1	1	50													50
aadA4, aac6Ib	Aminoglycoside resistant	25	CN	22	3	88	TOB	25	0	100													94
aadB, aadA1	Aminoglycoside resistant	1	CN	0	1	0	TOB	0	1	0													0
aphA	Aminoglycoside resistant	6	CN	0	6	100	TOB	0	6	100													100
aphA, aadA1	Aminoglycoside resistant	4	CN	0	4	100	TOB	0	4	100													100
aphA, aadA1, aadA2	Aminoglycoside resistant	6	CN	6	0	100	TOB	0	6	0													50
aphA, aadA1, aadA4	Aminoglycoside resistant	1	CN	1	0	100	TOB	1	0	100													100
aphA, aadA1, aadA4, aac6Ib, aac3Va	Aminoglycoside resistant	1	CN	1	0	100	TOB	1	0	100													100
aphA, aadA1, aadA2, aadA4	Aminoglycoside resistant	1	CN	1	0	100	TOB	1	0	100													100
n.d.	Aminoglycoside susceptible	29	CN	9	20	69	TOB	9	20	69													69

(Continued)

TABLE 2 | Continued

AMR genotype detected by microarray ^a	Expected AMR phenotype	Frequency of genotype detected by microarray	Detected phenotype												Concordance of expected AMR phenotype (%)
			Antibiotic tested						Concordance (%)						
			Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)		
		114	68	46	73	65	49	67							77
tetA	Tetracycline resistant	55	TE	55	0	100	TGC	1	54	98					99
tetA, tetB	Tetracycline resistant	14	TE	14	0	100	TGC	0	14	100					100
tetB	Tetracycline resistant	26	TE	26	0	100	TGC	0	26	100					100
tetD	Tetracycline resistant	6	TE	6	0	100	TGC	0	6	100					100
n.d.	Tetracycline susceptible	13	TE	4	9	0	TGC	0	13	100					50
		114	105	9	80	1	113	100							90
qepA	Fluoroquinolone resistant	14	CIP	14	0	100	MXF	14	0	100					100
qepA, qnrA1	Fluoroquinolone resistant	2	CIP	2	0	100	MXF	2	0	100					100
qnrA1	Fluoroquinolone resistant	31	CIP	30	1	97	MXF	30	1	97					97
qnrB	Fluoroquinolone resistant	1	CIP	1	0	100	MXF	1	0	100					100
qnrS	Fluoroquinolone resistant	10	CIP	3	7	30	MXF	8	2	80					55
qnrS, qnrA1	Fluoroquinolone resistant	4	CIP	2	2	50	MXF	3	1	75					63
n.d.	Fluoroquinolone susceptible	52	CIP	34	18	35	MXF	34	18	35					35
		114	86	28	73	92	22	84							79
(Continued)															

TABLE 2 | Continued

AMR genotype detected by microarray ^a	Expected AMR phenotype	Frequency of genotype detected by microarray	Detected phenotype										Concordance of expected AMR phenotype (%)	
			Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)	Antibiotic tested	Resistant	Susceptible	Concordance (%)
sul1, dfrA17	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul2, dfrA1	Co-trimoxazol resistant	6	STX	6	0	100								100
sul1, sul2, dfrA17	Co-trimoxazol resistant	22	STX	22	0	100								100
sul1, sul2, dfrA17, dfrA19	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul2, dfrA5	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul2, sul3, dfrA12	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul2, sul3, dfrA12, dfrA17	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul2, sul3, dfrA17	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul3, dfrA5, dfrA7, dfrA14, dfrA15, dfrA17	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul3, dfrA7	Co-trimoxazol resistant	1	STX	1	0	100								100
sul1, sul3, dfrA7, dfrA17	Co-trimoxazol resistant	1	STX	1	0	100								100
sul2, dfrA1	Co-trimoxazol resistant	7	STX	7	0	100								100
sul2, dfrA12, dfrA17	Co-trimoxazol resistant	1	STX	1	0	100								100
sul2, dfrA14	Co-trimoxazol resistant	14	STX	14	0	100								100
sul2, dfrA14, dfrA15	Co-trimoxazol resistant	1	STX	1	0	100								100
sul2, dfrA17	Co-trimoxazol resistant	4	STX	4	0	100								100

(Continued)

TABLE 2 | Continued

AMR genotype detected by microarray ^a	Expected AMR phenotype	Frequency of genotype detected by microarray	Detected phenotype						Concordance of expected AMR phenotype (%)				
			Antibiotic tested			Concordance (%)							
			Antibiotic tested	Resistant	Susceptible	Antibiotic tested	Resistant	Susceptible	Antibiotic tested	Resistant	Susceptible		
sul2, dfrA5 sul2, sul3, dfrA14 sul3, dfrA14	Co-trimoxazol resistant	2	STX	2	0	100							100
	Co-trimoxazol resistant	2	STX	2	0	100							100
	Co-trimoxazol resistant	1	STX	1	0	100							100
sul1 sul2	Co-trimoxazol susceptible	1	STX	1	0	0							0
	Co-trimoxazol susceptible	3	STX	1	2	66							66
	Co-trimoxazol susceptible	30	STX	29	1	3							3
sul1, sul2 sul1, sul2, sul3	Co-trimoxazol susceptible	1	STX	1	0	0							0
	Co-trimoxazol susceptible	1	STX	1	0	0							0
	Co-trimoxazol susceptible	1	STX	1	0	0							0
dfrA14 dfrA17 n.d.	Co-trimoxazol susceptible	8	STX	0	8	100							100
114			103			11	80			80			

Overall Concordance (%)83.5

^a A genotype was defined as a group of genes which have been described to confer resistance to a family of antibiotics (e.g., the genotype "aac6, aac6lb, aadA1" confers resistance to aminoglycosides). The phenotype was detected by a VITEK-2 system.

^b AK amikacin, CN gentamicin, TOB tobramycin, FEP cefepime, CTX cefotaxime, CAZ ceftazidime, MEM meropenem, IMP imipenem, TE tetracycline, TGC tigecycline, STX co-trimoxazole, CIP ciprofloxacin, MXF moxifloxacin.

^c 3G cephalosporin – 3rd generation cephalosporin.

Sample source	Sampling date	Farm ID	Isolate	O-type											H-type																	
				O17/O17											O17/O17																	
				1	4	8	9	53	83	86	99	101	112	114	138	2	4	5	6	9	10	11	18	20	21	25	30	31	32	34	42	45
1	HDC1	D	Jan 14	CS1																												
2	HDC2	D	Jan 14	CS3																												
3	HDC11	D	Jan 14	CS11																												
4	HDC18	D	Jan 14	CS20																												
5	HDC12	D	Jan 14	CS12																												
6	HDC17	D	Jan 14	CS19																												
7	HDC22	D	Jan 14	CS27																												
8	ES55	GF1	Feb 14	WT																												
9	HDC171	GF2	May 14	CS50																												
10	HDC64	GF2	Feb 14	CS40																												
11	HDC69	GF2	Feb 14	CS45																												
12	HDC59_KC	GF2	Feb 14	CS35																												
13	HDC70	GF2	Feb 14	CS47																												
14	HDC71	GF2	Feb 14	CS48																												
15	HDC75	GF2	Feb 14	CS53																												
16	HDC102	D	Apr 14	CS2																												
17	HDC107	D	Apr 14	CS8																												
18	HDC78	GF2	Feb 14	CS56																												
19	HDC56	GF2	Feb 14	CS31																												
20	HDC63	GF2	Feb 14	CS39																												
21	HDC57	GF2	Feb 14	CS32																												
22	HDC58	GF2	Feb 14	CS34																												
23	HDC60	GF2	Feb 14	CS36																												
24	HDC72	GF2	Feb 14	CS49																												
25	HDC91	GF5	Mar 14</																													

prevalence of ESBL-producing bacteria in livestock of healthy animals (Schmid et al., 2013; Dahms et al., 2015). Like in the present study, in both reports the most prevalent ESBL group was CTX-M. Schmid et al. (2013) collected a total of 598 samples that yielded 196 ESBL-producing *E. coli* (32.8%). The high percentage of ESBL-producing *E. coli* in healthy animals shows the high zoonotic risk for people working in close contact to animals. With this background Dahms et al. (2015) investigated different farms for ESBL-producing bacteria in samples collected from livestock and also from farm workers. In total, 70.6% of the tested farms and 5.8% of the farm workers were positive for ESBL-producing bacteria. In contrast, a study from Burgundy in France in 2012 showed only a low prevalence, of about 5%, of ESBL-producing bacteria in feces samples from different farms (Hartmann et al., 2012).

Due to the preselection of all samples on Brilliance™ ESBL agar, carbapenem resistant isolates were also retrieved. In six isolates, carbapenemase genes were detected using the CarbDetect AS-1 kit. In five isolates from a stall on farm GF1, *blaOXA-48* was found. These isolates were phenotypically and genotypically identical. Interestingly, another isolate was found containing *blaOXA-181*, a gene belonging to the *blaOXA-48*-like family. Both genes are also widely distributed in human pathogens that cause hospital-acquired infection (Poirel et al., 2012). Carbapenems are last-line antibiotics with a broad spectrum and a high efficacy, and are stable against ESBLs. As they can only be administered intravenously, carbapenems are used exclusively in the clinical environment, and are not known to be used in animal husbandry. While carbapenemase-producing *Enterobacteriaceae* (CPE) are mostly described in hospitalized humans (Nordmann et al., 2011; Abdallah et al., 2015b), the current study shows conclusively that such CREs are also found in farm animals including apparently healthy dairy cattle. The detection of carbapenem resistant isolates in such an environment and the threat that such multi-drug resistance bacteria ends up in consumer food (e.g., milk or dairy products), raises serious concerns about public health. Carbapenemase-producing isolates have been detected in poultry farms (Abdallah et al., 2015a), but there are to date no reports describing finding of these multi-drug resistant bacteria in dairy cattle farms in Egypt.

Given that the majority of samples that contained ESBL/carbapenemase-producing bacteria originated from rectal swabs, this raises the question of how the potentially contaminated feces are disposed of. Normally, dung is used as fertilizer in agriculture. Via this route, multi-drug resistant pathogens might get into the food chain, either directly through

consumption of meat, or indirectly from cattle grazing on fertilized pasture. Another major problem raises up in this context, the resistance genes described in this paper are usually found on plasmids (Chanawong et al., 2001; Paterson and Bonomo, 2005; Duan et al., 2006; Wittum et al., 2010; Brolund, 2014; Hansen et al., 2014; Valentin et al., 2014) and such mobile elements can be easily transferred to and between environmental bacteria (Aminov, 2011; Berglund, 2015), as well as to other human pathogens (Pitout et al., 2015). This poses a high risk to the environment, and the human population.

CONCLUSION

To the best of our knowledge, this study is the first report which analyses the prevalence of ESBL-producing *E. coli* in Egyptian dairy farms based on genotyping and phenotyping data. The high percentage of ESBL-positive isolates and even of carbapenemase-producing bacteria was alarming given the relevance of 3G cephalosporins and carbapenems in modern medicine. Additionally, the isolates had a highly diverse genetic background with regard to serotype, virulence and antimicrobial resistance markers (Figures 2, 3). Experiments showed a high degree of concordance between genotype and phenotype.

Strict hygiene measures are mandatory to control the spread, the transmission dynamics and potential zoonotic risk factors of ESBL- and carbapenemase-producing bacteria in dairy farms.

AUTHOR CONTRIBUTIONS

SB, HE, and HH conceived of the study, and participated in its design and coordination. SB, IE, and DW carried out the genotyping and serotyping. SB and IE carried out the antimicrobial resistance pattern by VITEK-2. MA and HE participated in sampling. HE, HH, and MA participated in preliminary design as well as bacteriological analysis of part of the study. SB, SM, HE, HH, and RE drafted the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Keri Clack (Alere Technologies, Jena, Germany) for proof reading the manuscript, Annett Reißig (Alere Technologies, Jena, Germany) and Elke Müller (Alere Technologies, Jena, Germany) for excellent technical support, Christina Braun for continuous support as well as for proof reading of the manuscript.

REFERENCES

- Abdallah, H. M., Reuland, E. A., Wintermans, B. B., Al Naiemi, N., Koek, A., Abdelwahab, A. M., et al. (2015a). Extended-spectrum beta-lactamases and/or carbapenemases-producing *Enterobacteriaceae* isolated from retail chicken meat in Zagazig, Egypt. *PLoS ONE* 10:e0136052. doi: 10.1371/journal.pone.0136052
- Abdallah, H. M., Wintermans, B. B., Reuland, E. A., Koek, A., al Naiemi, N., Ammar, A. M., et al. (2015b). Extended-spectrum beta-lactamase- and carbapenemase-producing *Enterobacteriaceae* isolated from Egyptian patients with suspected blood stream infection. *PLoS ONE* 10:e0128120. doi: 10.1371/journal.pone.0128120
- Abraham, S., Chapman, T. A., Zhang, R., Chin, J., Mabbett, A. N., Totsika, M., et al. (2012). Molecular characterization of *Escherichia coli* strains that cause symptomatic and asymptomatic urinary tract infections. *J. Clin. Microbiol.* 50, 1027–1030. doi: 10.1128/JCM.06671-11
- Aminov, R. I. (2011). Horizontal gene exchange in environmental microbiota. *Front. Microbiol.* 2:158. doi: 10.3389/fmicb.2011.00158

- Berglund, B. (2015). Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics. *Infect. Ecol. Epidemiol.* 5:28564. doi: 10.3402/iee.v5.28564
- Bradford, P. A. (2001). Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clin. Microbiol. Rev.* 14, 933–951. doi: 10.1128/CMR.14.4.933-951.2001
- Braun, S. D., Monecke, S., Thurmer, A., Ruppelt, A., Makarewicz, O., Pletz, M., et al. (2014). Rapid identification of carbapenemase genes in gram-negative bacteria with an oligonucleotide microarray-based assay. *PLoS ONE* 9:e102232. doi: 10.1371/journal.pone.0102232
- Brinas, L., Moreno, M. A., Zarazaga, M., Porrero, C., Saenz, Y., Garcia, M., et al. (2003). Detection of CMY-2, CTX-M-14, and SHV-12 beta-lactamases in *Escherichia coli* fecal-sample isolates from healthy chickens. *Antimicrob. Agents Chemother.* 47, 2056–2058. doi: 10.1128/AAC.47.6.2056-2058.2003
- Brolund, A. (2014). Overview of ESBL-producing *Enterobacteriaceae* from a Nordic perspective. *Infect. Ecol. Epidemiol.* 4, 1–9. doi: 10.3402/iee.v4.24555
- Carattoli, A. (2008). Animal reservoirs for extended spectrum beta-lactamase producers. *Clin. Microbiol. Infect.* 14(Suppl. 1), 117–123. doi: 10.1111/j.1469-0691.2007.01851.x
- Chanawong, A., M'Zali, F. H., Heritage, J., Lulitanond, A., and Hawkey, P. M. (2001). SHV-12, SHV-5, SHV-2a and VEB-1 extended-spectrum beta-lactamases in Gram-negative bacteria isolated in a university hospital in Thailand. *J. Antimicrob. Chemother.* 48, 839–852. doi: 10.1093/jac/48.6.839
- Dahms, C., Hubner, N. O., Kossow, A., Mellmann, A., Dittmann, K., and Kramer, A. (2015). Occurrence of ESBL-producing *Escherichia coli* in livestock and farm workers in Mecklenburg-Western Pomerania, Germany. *PLoS ONE* 10:e0143326. doi: 10.1371/journal.pone.0143326
- Dahshan, H., Abd-Elall, A. M., Megahed, A. M., Abd-El-Kader, M. A., and Nabawy, E. E. (2015). Veterinary antibiotic resistance, residues, and ecological risks in environmental samples obtained from poultry farms, Egypt. *Environ. Monit. Assess.* 187, 1–10. doi: 10.1007/s10661-014-4218-3
- Duan, R. S., Sit, T. H., Wong, S. S., Wong, R. C., Chow, K. H., Mak, G. C., et al. (2006). *Escherichia coli* producing CTX-M beta-lactamases in food animals in Hong Kong. *Microb. Drug Resist.* 12, 145–148. doi: 10.1089/mdr.2006.12.145
- Giamarellou, H. (2005). Multidrug resistance in Gram-negative bacteria that produce extended-spectrum beta-lactamases (ESBLs). *Clin. Microbiol. Infect.* 11(Suppl. 4), 1–16. doi: 10.1111/j.1469-0691.2005.01160.x
- Haftu, R., Taddele, H., Gugsu, G., and Kalayou, S. (2012). Prevalence, bacterial causes, and antimicrobial susceptibility profile of mastitis isolates from cows in large-scale dairy farms of Northern Ethiopia. *Trop. Anim. Health Prod.* 44, 1765–1771. doi: 10.1007/s11250-012-0135-z
- Hansen, K. H., Bortolaia, V., Damborg, P., and Guardabassi, L. (2014). Strain diversity of CTX-M-producing *Enterobacteriaceae* in individual pigs: insights into the dynamics of shedding during the production cycle. *Appl. Environ. Microbiol.* 80, 6620–6626. doi: 10.1128/AEM.01730-14
- Hartmann, A., Locatelli, A., Amoureux, L., Depret, G., Jolivet, C., Gueneau, E., et al. (2012). Occurrence of CTX-M producing *Escherichia coli* in soils, cattle, and farm environment in France (Burgundy Region). *Front. Microbiol.* 3:83. doi: 10.3389/fmicb.2012.00083
- ISO (2001). *ISO 16654:2001 Microbiology of Food and Animal Feeding Stuffs – Horizontal Method for the Detection of Escherichia coli O157* [Online]. Geneva. Available online at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=29821 [Accessed].
- Li, Y., Zheng, B., Zhu, S., Xue, F., and Liu, J. (2015). Antimicrobial susceptibility and molecular mechanisms of fosfomycin resistance in clinical *Escherichia coli* isolates in mainland China. *PLoS ONE* 10:e0135269. doi: 10.1371/journal.pone.0135269
- Nordmann, P., Naas, T., and Poirel, L. (2011). Global spread of carbapenemase-producing *Enterobacteriaceae*. *Emerging Infect. Dis.* 17, 1791–1798. doi: 10.3201/eid1710.110655
- Paterson, D. L., and Bonomo, R. A. (2005). Extended-spectrum beta-lactamases: a clinical update. *Clin. Microbiol. Rev.* 18, 657–686. doi: 10.1128/CMR.18.4.657-686.2005
- Pitout, J. D., Nordmann, P., and Poirel, L. (2015). Carbapenemase-producing *Klebsiella pneumoniae*, a key pathogen set for global nosocomial dominance. *Antimicrob. Agents Chemother.* 59, 5873–5884. doi: 10.1128/AAC.01019-15
- Poirel, L., Potron, A., and Nordmann, P. (2012). OXA-48-like carbapenemases: the phantom menace. *J. Antimicrob. Chemother.* 67, 1597–1606. doi: 10.1093/jac/dks121
- Ramirez, M. S., and Tolmasey, M. E. (2010). Aminoglycoside modifying enzymes. *Drug Resist. Updat.* 13, 151–171. doi: 10.1016/j.drug.2010.08.003
- Rawat, D., and Nair, D. (2010). Extended-spectrum beta-lactamases in gram negative bacteria. *J. Glob. Infect. Dis.* 2, 263–274. doi: 10.4103/0974-777X.68531
- Schmid, A., Hörmansdorfer, S., Messelhauser, U., Käsbohrer, A., Sauter-Louis, C., and Mansfeld, R. (2013). Prevalence of extended-spectrum beta-lactamase-producing *Escherichia coli* on Bavarian dairy and beef cattle farms. *Appl. Environ. Microbiol.* 79, 3027–3032. doi: 10.1128/AEM.00204-13
- Shaikh, S., Fatima, J., Shakil, S., Rizvi, S. M., and Kamal, M. A. (2015). Antibiotic resistance and extended spectrum beta-lactamases: types, epidemiology and treatment. *Saudi J. Biol. Sci.* 22, 90–101. doi: 10.1016/j.sjbs.2014.08.002
- Takahata, S., Ida, T., Hiraishi, T., Sakakibara, S., Maebashi, K., Terada, S., et al. (2010). Molecular mechanisms of fosfomycin resistance in clinical isolates of *Escherichia coli*. *Int. J. Antimicrob. Agents* 35, 333–337. doi: 10.1016/j.ijantimicag.2009.11.011
- Valentin, L., Sharp, H., Hille, K., Seibt, U., Fischer, J., Pfeifer, Y., et al. (2014). Subgrouping of ESBL-producing *Escherichia coli* from animal and human sources: an approach to quantify the distribution of ESBL types between different reservoirs. *Int. J. Med. Microbiol.* 304, 805–816. doi: 10.1016/j.ijmm.2014.07.015
- WHO (2013). *Consultative Meeting on Antimicrobial Resistance for Countries in the Eastern Mediterranean Region: From Policies to Action* [Online]. Sharm El Sheikh. Available online at: http://applications.emro.who.int/docs/IC_Meet_Rep_2014_EN_15210.pdf [Accessed 2013].
- Wittum, T. E., Mollenkopf, D. F., Daniels, J. B., Parkinson, A. E., Mathews, J. L., Fry, P. R., et al. (2010). CTX-M-type extended-spectrum beta-lactamases present in *Escherichia coli* from the feces of cattle in Ohio, United States. *Foodborne Pathog. Dis.* 7, 1575–1579. doi: 10.1089/fpd.2010.0615

Conflict of Interest Statement: SB, DW, SM, IE, and RE are employees of Alere Technologies GmbH, the company that manufactures the microarrays also used in this study. This has no influence on study design, data collection and analysis, and this does not alter the authors' adherence to all the Frontiers policies on sharing data and materials.

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

Copyright © 2016 Braun, Ahmed, El-Adawy, Hotzel, Engelmann, Weiß, Monecke and Ehrlich. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Development of a High Resolution Virulence Allelic Profiling (HReVAP) Approach Based on the Accessory Genome of *Escherichia coli* to Characterize Shiga-Toxin Producing *E. coli* (STEC)

Valeria Michelacci^{1*}, Massimiliano Orsini², Arnold Knijn³, Sabine Delannoy⁴, Patrick Fach⁴, Alfredo Caprioli¹ and Stefano Morabito¹

OPEN ACCESS

Edited by:

Pina Fratamico,
United States Department of
Agriculture-Agricultural Research
Service, USA

Reviewed by:

James L. Bono,
United State Department of
Agriculture- Agricultural Research
Service, USA
Erin R. Reichenberger,
United States Department of
Agriculture, USA

*Correspondence:

Valeria Michelacci
valeria.michelacci@iss.it

Specialty section:

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

Received: 26 November 2015

Accepted: 05 February 2016

Published: 23 February 2016

Citation:

Michelacci V, Orsini M, Knijn A,
Delannoy S, Fach P, Caprioli A and
Morabito S (2016) Development of a
High Resolution Virulence Allelic
Profiling (HReVAP) Approach Based
on the Accessory Genome of
Escherichia coli to Characterize
Shiga-Toxin Producing *E. coli* (STEC).
Front. Microbiol. 7:202.
doi: 10.3389/fmicb.2016.00202

¹ European Reference Laboratory for *Escherichia coli*, Dipartimento di Sanità Pubblica Veterinaria e Sicurezza Alimentare, Istituto Superiore di Sanità, Rome, Italy, ² Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise G. Caporale, Teramo, Italy, ³ Servizio Informatico, Documentazione, Biblioteca e Attività Editoriali, Istituto Superiore di Sanità, Rome, Italy, ⁴ Platform IdentityPath, Food Safety Laboratory, ANSES, Université Paris-Est, Maisons-Alfort, France

Shiga-toxin producing *Escherichia coli* (STEC) strains possess a large accessory genome composed of virulence genes existing in multiple allelic variants, which sometimes segregate with specific STEC subpopulations. We analyzed the allelic variability of 91 virulence genes of STEC by Real Time PCR followed by melting curves analysis in 713 *E. coli* strains including 358 STEC. The 91 genes investigated were located on the locus of enterocyte effacement (LEE), OI-57, and OI-122 pathogenicity islands and displayed a total of 476 alleles in the study population. The combinations of the 91 alleles of each strain were termed allelic signatures and used to perform cluster analyses. We termed such an approach High Resolution Virulence Allelic Profiling (HReVAP) and used it to investigate the phylogeny of STEC of multiple serogroups. The dendrograms obtained identified groups of STEC segregating approximately with the serogroups and allowed the identification of subpopulations within the single groups. The study of the allelic signatures provided further evidence of the coevolution of the LEE and OI-122, reflecting the occurrence of their acquisition through a single event. The HReVAP analysis represents a sensitive tool for studying the evolution of LEE-positive STEC.

Keywords: accessory genome, allelic variants, STEC subtyping, phylogenesis, bioinformatics

INTRODUCTION

Human infections with Shiga-toxin producing *Escherichia coli* (STEC) cause a wide range of symptoms including uncomplicated diarrhea, hemorrhagic colitis, and the life-threatening hemolytic uremic syndrome (HUS) (Caprioli et al., 2005). The main virulence feature of STEC is the ability to produce Shiga-toxins (Stx), which interfere with the protein synthesis in the target cells, eventually causing their death (O'Brien and Holmes, 1987). The capacity to produce Stx is acquired through infection with bacteriophages conveying the *stx* genes, which can remain stably integrated into the bacterial chromosome (O'Brien et al., 1984).

In spite of the striking biological effect exerted by the Stx, their sole production seems not to be sufficient for causing the disease, at least the most severe forms. As a matter of fact, only a few STEC serogroups are usually isolated from human cases of severe disease (Nataro and Kaper, 1998; Karmali et al., 2003), which share the presence in the genome of mobile genetic elements (MGEs) encoding robust machineries for the colonization of the host gut (McDaniel and Kaper, 1997; Paton et al., 2001; Morabito et al., 2003; Imamovic et al., 2010; Michelacci et al., 2013). Three Pathogenicity Islands (PAIs) have been described in the genome of such STEC serogroups: the *locus of enterocyte effacement* (LEE) (McDaniel and Kaper, 1997), the OI-122 (Karmali et al., 2003; Morabito et al., 2003), and the OI-57 (Imamovic et al., 2010).

The LEE locus governs the ability to induce the typical “attachment and effacement” (A/E) lesion on the enterocyte. It encodes a type three secretion system, effectors subverting the cell functions related with the cytoskeleton assembly and maintenance, and factors mediating the intimate adhesion of the bacterium to the enterocyte, including the adhesin intimin (McDaniel and Kaper, 1997). The other two PAIs carry genes whose products are also involved in the mechanism of colonization, such as Efa1/LifA, encoded by a gene present in the OI-122 (Morabito et al., 2003), and AdfO (Ho et al., 2008), whose genetic determinant is conveyed by the OI-57 (Imamovic et al., 2010).

During the last decades different authors deployed schemes for the classification of the different STEC types (Griffin and Tauxe, 1991; Nataro and Kaper, 1998; Karmali et al., 2003). One of these schemes groups the STEC strains based on the serogroup, relative incidence of human infections, ability to cause severe diseases, association with outbreaks and presence of virulence-associated MGEs in the genome (Karmali et al., 2003). According to this classification, STEC are divided into seropathotypes (SPTs), identified with letters from A to E in a decreasing rank of pathogenicity. SPT A comprises STEC O157, while SPT B includes the STEC belonging to serogroups different from O157 but causing both sporadic cases and outbreaks of HUS, namely O26, O103, O111, O145, and O121. SPTs A and B share the presence of the LEE, OI-57, and OI-122 PAIs in their genome. The SPT C includes a number of STEC serogroups, including O113 and O91, which apparently do not harbor the LEE locus but are sporadically isolated from severe infections. Finally, STEC included in the SPTs D and E have rarely or never been associated with human disease respectively (Karmali et al., 2003). For the last three SPTs the information on the presence and integrity of the three PAIs are scanty.

The complexity of the STEC virulome is an important source of strain genomic variability, which is further augmented by the existence of multiple allelic variants of the virulence genes. Some of the subtypes of *stx2* have been significantly associated with the most severe infection (Friedrich et al., 2002), while some other subtypes of both *stx1* and *stx2* seemed to be primarily associated with a milder course of the disease or confined to animal hosts (Friedrich et al., 2002; Bielaszewska et al., 2006; Persson et al., 2007; Scheut et al., 2012). A considerable heterogeneity has also been identified in the DNA sequence of the intimin-coding gene *eae*, leading to the identification of at least 18 intimin types

unevenly distributed in the different STEC serogroups (Oswald et al., 2000; Tarr et al., 2002; Ito et al., 2007; Madic et al., 2010).

In the present study we developed an approach to simultaneously identify the presence and the allelic types of a large panel of genes carried by the LEE locus, OI-122, and OI-57 PAIs and used it to study the phylogeny of STEC belonging to SPT A, B and C.

MATERIALS AND METHODS

Bacterial Strains

A total of 713 *E. coli* strains positive for at least one of the three pathogenicity islands LEE, OI-122, and OI-57 were selected among the isolates present in the culture collections of the Istituto Superiore di Sanità (ISS, Rome, Italy) and the Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du travail (ANSES, Maisons Alfort, France) and used to identify the alleles of the genes harbored by the three PAIs. The panel comprised 358 STEC strains belonging to serogroups O157 ($n = 81$), O26 ($n = 32$), O111 ($n = 36$), O103 ($n = 8$), O145 ($n = 8$), O121 ($n = 3$), and others ($n = 190$) isolated from unrelated human cases of human infections and from food in Italy and France in the period 2008–2011. Additional 355 *stx*-negative *E. coli* of multiple serogroups and isolated from human, food and animal sources in the same countries and period were included in the study. The O157 STEC strain EDL933 was used as positive control (Supplementary Table 1, Sheet 1).

A population of 318 unrelated STEC strains, part of the panel of strains described above, was used to assess the performance of the HReVAP approach. These included 161 isolates of SPTs A and B and belonging to serogroups O157 ($n = 81$), O26 ($n = 32$), O111 ($n = 33$), O103 ($n = 5$), O145 ($n = 7$) and O121 ($n = 3$) and 157 *eae*-negative strains, of serogroups O91 ($n = 14$), O174 ($n = 10$), O113 ($n = 9$), O104 ($n = 6$), O101 ($n = 3$), O153 ($n = 3$), O21 ($n = 3$), and others ($n = 109$). The study population also included a panel of 36 *stx*-negative *eae*-positive *E. coli*, altogether referred to as EPEC, including the following serogroups: O26 ($n = 12$), O127 ($n = 5$), O55 ($n = 4$), O128 ($n = 3$), O125 ($n = 3$), O111 ($n = 2$), O86 ($n = 2$), and others ($n = 5$).

Finally, 39 out of the 161 SPTs A and B STEC were also subjected to whole genome sequencing, followed by Multi Locus Sequence Typing (MLST) and Whole genome SNP analysis with the aim of comparing HReVAP results with those obtained using these DNA-sequence-based methods. The latter isolates included strains belonging to serogroups O157 ($n = 16$), O26 ($n = 12$), O111 ($n = 7$), O103 ($n = 2$), O145 ($n = 1$), and O121 ($n = 1$).

Real-Time PCR and Melting Curves Analysis

Ninety-one primer pairs were deployed to amplify 100–300 bp fragments from as many genes harbored on the three PAIs LEE, OI-122, and OI-57 (38, 12, and 41 genes from each island, respectively). The genomic sequence of the O157 STEC strain EDL933 (Acc. no. AE005174) was used to design the primer pairs using the Primer-BLAST web-tool available on the NCBI webserver. Some of the primers were degenerated to amplify the

target genes in all the STEC strains for which a sequence was available in GenBank. The sequences of primers used in this study and their annealing position on the genomic sequence of EDL933 strain are reported in **Table 1**.

Total DNA was extracted from overnight cultures of the strains with the Nucleospin Tissue extraction kit (Macherey-Nagel, Düren, DE).

The Real-Time PCR reactions were performed on the high throughput BioMark Real Time PCR system with 96.96 Genotyping Dynamic Array Chips (Fluidigm, San Francisco, CA), using the EvaGreen DNA binding dye (Biotium Inc., Hayward, CA). The thermal profile was 95°C for 10 min (enzyme activation) followed by 35 cycles of 95°C for 15 s and 60°C for 1 min (amplification step). Finally, a denaturation step was performed and the melting curves of the amplified products were registered. Eight array chips were used to perform the whole panel of reactions, each including a positive template control consisting in the DNA extracted from an overnight culture of the EDL933 strain. In addition to the 91 genes, each sample was also subjected to amplification of the *stx* genes (Perelle et al., 2004) and the *wecA* housekeeping gene, used as marker for *E. coli* species (Forward primer: 5'-CTTTATCTCAGTAGCCTGGG-3', Reverse primer: 5'-AGGAAGTAACCAAACGGTCC-3').

High Resolution Virulence Allelic Profiling Analysis (HReVAP)

The melting temperatures (T_m) of the amplicons were normalized using the T_m of the PCR products amplified from the positive control strain EDL933 present in each array chip. Normalization values were obtained for each array chip minimizing the mean of the T_m standard deviations for the eight normalized T_m measurements over each gene as well as the overall maximum value of the normalized T_m ranges of all genes.

The T_m frequency distributions for each gene have been calculated by grouping values in 0.05°C intervals. The frequency distributions were then analyzed with the “mix” function of the “mixdist” package in the R software (RTeam, 2014). This function finds Maximum Likelihood estimates for the proportions, means, and standard deviations of a mixture distribution by applying a Newton-type iterative method (RTeam, 2014). The number of Gaussians and the starting parameters were adjusted upon evaluation of several fitting results. In order to limit the degrees of freedom, prior knowledge was applied to the model: in each model fit, the standard deviations (σ) of the Gaussian curves were left variable but constrained to be equal. The following command was used for the analysis: `mix[Tm_dat, Tm_par, dist = “norm,” mixconstr(conmu = “NONE,” consigma = “SEQ”)]`, where `Tm_dat` is the data frame of the grouped T_m data and `Tm_par` a data frame of the starting values for the parameters of the distributions. Several normal distributions were observed for each gene at different intervals of temperatures. Before clustering, T_m values were aggregated into numbered classes utilizing the model fitting results to define temperature intervals for distinct alleles. The interval limits were calculated as the points of intersection between two adjoining Gaussian curves with the

same standard deviation according to the following equation:

$$T_{int} (^{\circ}\text{C}) = \frac{\mu_1^2 - \mu_2^2 - 2\sigma^2 \ln\left(\frac{\pi_1}{\pi_2}\right)}{2\mu_1 - 2\mu_2} \quad (1)$$

T_{int} = Temperature at the point of intersection between Gaussian 1 and 2

μ_i = Mean temperature of the Gaussian curve $i = 1, 2$

σ = Standard deviation of the Gaussian curves

π_i = Amplitude of the Gaussian curve $i = 1, 2$

The model fitting used did not prove optimal for allele assignment due to the proximity of many Gaussian curves, with consequent overlap. Therefore, all fits have been revised trying to keep the maximum number of Gaussian curves in the model while aggregating those heavily overlapping. With this procedure, the resolution of the method (capacity of allele distinction) has been lowered in benefit of the precision of the method (convergence of assignments). The T_m intervals identified by peaks in the distributions obtained from the analysis of each gene were used to identify the alleles, which were labeled with numbers in ascending order according to the position of the peak in the temperature distribution (Supplementary Table 2). Each strain was given a numeric allelic signature comprising the alleles of all the genes analyzed.

The *neighbor* software of the Emboss package for samples clustering with default parameters (Rice et al., 2000) was used to compare the allelic signatures and to obtain distance matrices.

Each step of the above described procedure has been implemented in dedicated python classes, including the *neighbor* software that was wrapped together with the TreeGraph software (Stover and Muller, 2010) for graphical trees representation.

The HReVAP software package was deployed and used on the public computational framework ARIES operating on the servers of the Istituto Superiore di Sanità and based on the Galaxy bioinformatics platform (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010) (<https://w3.iss.it/site/aries/>).

The tree files (.tre) produced with the HReVAP clustering algorithm were downloaded and visualized using the FigTree program version 1.4.0 (Drummond et al., 2012).

Whole Genome Sequencing of STEC and Phylogenetic Analyses

Thirty-nine out of the 161 STEC strains used for the HReVAP typing were subjected to whole genome sequencing using the Library Preparation Kit by Kapa Biosystems (Wilmington, MA, USA) and a paired end 100 bp protocol on an Illumina HiSeq2500 instrument in fast run mode according to manufacturers' instructions. The sequencing reads have been uploaded in the EMBL-ENA sequence database (EMBL European Nucleotide Archive Study accession no. PRJEB11886). The raw reads were trimmed to remove the adaptors and to accept 27 as the lowest Phred value and assembled using the *de novo* assembly tool Edena v3 (Hernandez et al., 2008). The contigs were subjected to *in silico* Multi Locus Sequence Typing (MLST) with the protocol described by Wirth and colleagues (Wirth et al., 2006). Whole

genome Single Nucleotide Polymorphism (WG-SNP) analysis was performed using the ksnp3 pipeline (Gardner et al., 2015), using 19 as kmer size. The optimum value for the kmer size was selected as that producing the highest number of unique kmers of the median length in all the genomes of the dataset and it was calculated by using the kchooser tool included in the ksnp3 pipeline. All the bioinformatics analyses were performed through the ARIES webserver (<https://w3.iss.it/site/aries/>).

RESULTS

HReVAP: Identification of the Alleles

The analysis of the 91 genes conveyed by the three pathogenicity islands LEE, OI-122, and OI-57 allowed identifying a total of 476 alleles (Supplementary Table 1, Sheet 1 and Supplementary Table 3). Each gene displayed 2 to 10 different alleles in the study population.

The *eae*-positive strains included in the panel exhibited the presence of 32 out of the 38 LEE-harbored genes on average, while the strains positive for the marker of OI-122, *efa1-lifA*, were positive for the majority of the 12 genes selected on this PAI (10.8 genes on average). The OI-57 showed the widest variability. As a matter of fact, 10.4% of the isolates proved positive for 1 to 10 genes of PAI OI-57, 30.6% were positive for 11–20 genes, 26% fell in the range of 21–30 genes detected and 33% gave positive result for more than 31 out of the 41 targets considered (Supplementary Table 1, Sheet 1).

The genes conveyed by the LEE and the OI-122 islands showed a mean number of alleles of 4.79 (range: 2–8 alleles; median = 5) and 4.17 (range: 3–8 alleles; median = 4) for each gene, respectively, while those part of the OI-57 were the most variable, displaying a mean value of 5.95 alleles each (range: 3–10 alleles; median = 6) (Figure 1 and Supplementary Table 3). Interestingly, the LEE locus displayed a uniform allelic variation throughout its whole length while the OI-122 and the OI-57 appeared to have a slightly higher number of alleles in the leftmost part (Figure 1).

HReVAP Performance: Amplification of the LEE, OI-122, and OI-57 Targets in STEC and EPEC

All the LEE-genes could be amplified in the vast majority of the STEC strains belonging to SPTs A and B (Figure 2A and Supplementary Table 1, Sheet 2). In detail, no negative results were obtained for all the O157 and O145 strains tested, with the only exception of one O145 strain, negative for eight targets. Nine LEE-borne genes, namely the open reading frames (ORF) Z5101, Z5107, Z5111, Z5112, Z5114, Z5117, Z5121, Z5122, and Z5127 were more variable in the STEC serogroups other than the O157 and O145, with more than 60% of the STEC O26 strains tested producing no amplicons. The same nine targets could not be amplified in all the STEC O103 and O121 strains, with a few exceptions (Figure 2A). Four of these nine gene targets (Z5107, Z5112, Z5114, and Z5122) could not be amplified from the whole

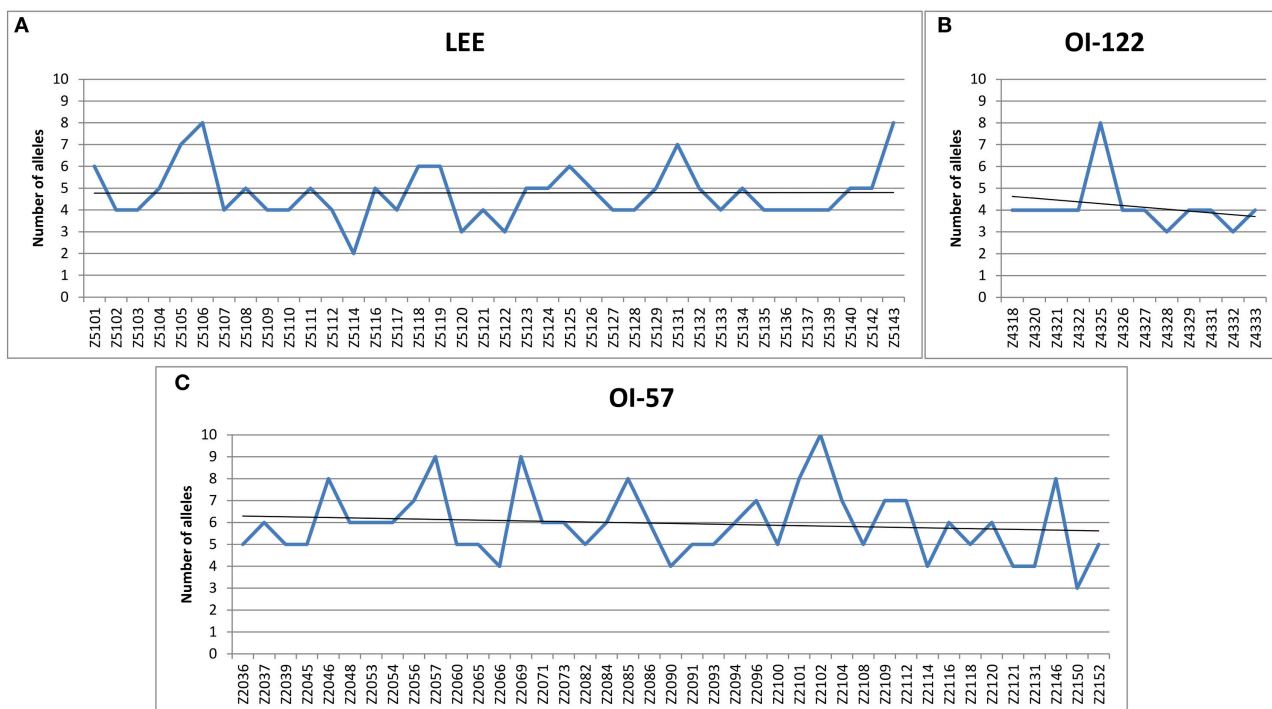
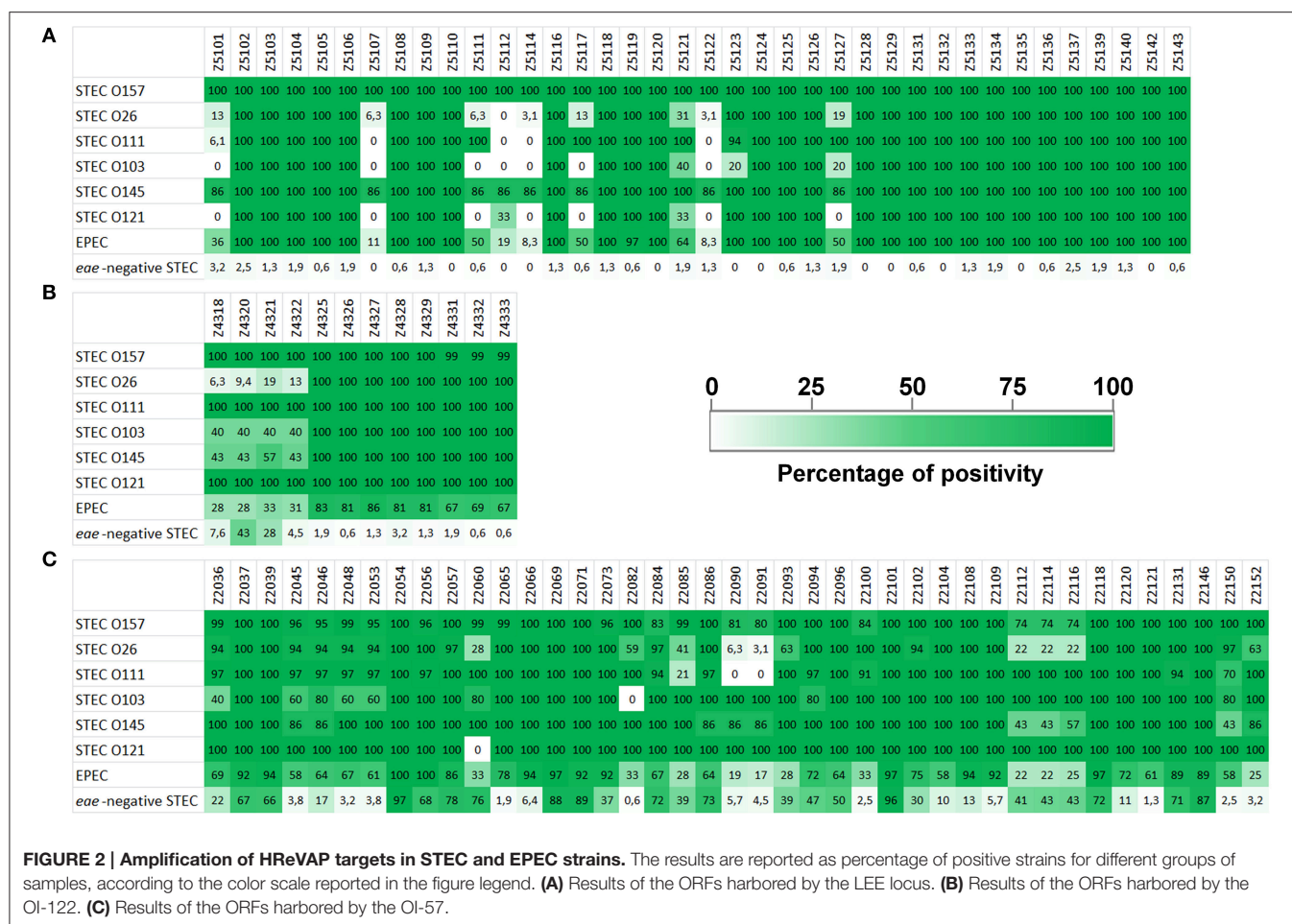


FIGURE 1 | Allelic variability of the PAIs assayed. Number of alleles identified for each of the ORFs assayed and trend lines. **(A)** Alleles identified in the ORFs harbored by the LEE locus. **(B)** Alleles identified in the ORFs harbored by the OI-122. **(C)** Alleles identified in the ORFs harbored by the OI-57.



panel of the STEC O111, while ORF Z5101 gave positive result in only two strains of this serogroup.

As for the OI-122 PAI, all the 12 ORFs selected were detected in the whole panel of STEC O157, O111, and O121 strains, with the only exception of one O157 strain, which was negative for the three ORFs Z4331, Z4332, and Z4333 (Figure 2B and Supplementary Table 1, Sheet 2).

The STEC strains belonging to serogroups O145, O103, and O26 showed different amplification profiles, with the four ORFs Z4318, Z4320, Z4321, and Z4322 negative in 60% of the O103 strains and in the majority of O145 and O26 strains (Figure 2B).

The HReVAP typing of the STEC and EPEC confirmed the highest degree of variation of PAI OI-57. In particular, ORFs Z2112, Z2114, and Z2116 could not be amplified in many STEC O26, O157, and O145 while ORFs Z2090 and Z2091 were not detected in all the STEC O111, in the majority of STEC O26 strains, and in some STEC O145 and O157. Finally, the ORF Z2085 was frequently negative in STEC O111 and O26 (Figure 2C and Supplementary Table 1, Sheet 2).

Unexpectedly, the *eae*-negative STEC strains tested also showed positivity to many targets of the PAI OI-57 (Figure 2 and Supplementary Table 1, Sheet 3). In particular, two ORFs, Z2054, and Z2101, were positive in more than 95% of the strains tested, while genes Z2037, Z2039, Z2056, Z2057, Z2060, Z2069, Z2071,

Z2084, Z2086, Z2096, Z2118, Z2131, and Z2146, were positive in more than 50% of the population assayed.

As a whole, a mean of 15.9 targets out of the 41 selected for OI-57 were present in the panel of *eae*-negative STEC strains (range: 5–39; median = 15) (Supplementary Table 1, Sheet 3).

The EPEC strains assayed provided different amplification patterns (Figure 2 and Supplementary Table 1, Sheet 4). As expected, the LEE-borne ORFs and the OI-122 targets followed a pattern of positivity to PCR similar to that displayed by the STEC belonging to SPTs A and B, although with some variation. The OI-57 was present in all the EPEC strains tested but showed two regions of major variability encompassing the ORFs Z2090–Z2093 (negativity range: 72.2–83.3%) and Z2112–Z2116 (negativity range: 75–77.78%) (Figure 2 and Supplementary Table 1, Sheet 4).

HReVAP Typing: Allelic Variability of the LEE, the OI-122, and the OI-57

The allelic variability of the genes harbored by the three pathogenicity islands LEE, OI-122, and OI-57 has been investigated in the same study population used to assess the HReVAP performance.

The clustering of the allelic signatures of the STEC strains of SPTs A and B produced a dendrogram whose branches segregated

with the serogroups, with a few exceptions (**Figure 3A**). In particular, the cluster formed by STEC O157 strains appeared clearly distinct from the others and much more homogeneous, while the strains belonging to O111 and O26 serogroups were divided in two and three distinct clusters, respectively. Similar results were obtained when the cluster analysis was carried out separately using the allelic signatures produced with the ORFs of the LEE locus (**Figure 3B**) and of the PAI OI-122 only (**Figure 3C**). Such dendrograms displayed the same topology of that produced when the alleles of the complete ORFs panel were used but showed a lower intra-cluster resolution. More complex results were instead obtained from the HReVAP analysis of the genes conveyed by the OI-57, reflecting the highest variability observed in the ORFs of this PAI (**Figure 3D**). Even if the main groups corresponding to STEC serogroups O157, O111, and O26 could still be detected, the overall topology showed wider and less defined branches.

The topology of the dendrograms obtained with the EPEC isolates resembled that of those produced with the STEC of SPTs A and B allelic signatures. However, the higher variability of

EPEC, together with the lower number of isolates tested, caused the output to be less definite (**Figure 4**).

As for the *eae*-negative STEC strains, the cluster analysis of the allelic signatures was carried out exploiting the observed positivity to many of the ORFs of the OI-57. Although based on a smaller number of targets, this analysis showed a massive variability in the allelic signatures, yet able to distinguish and group different populations of strains (**Figure 5**).

Comparison between HReVAP and the DNA Sequence-Based MLST and Whole Genome-SNP Analyses

Thirty-nine STEC strains were used to compare the results of the HReVAP with those produced by DNA sequence-based typing techniques. The strains were either analyzed with the HReVAP or their genomes subjected to *in silico* MLST and WG-SNP analysis. The comparison showed that the HReVAP (**Figure 6A**) had a much higher discriminatory power than the MLST (**Figure 6B**) and produced a dendrogram similar

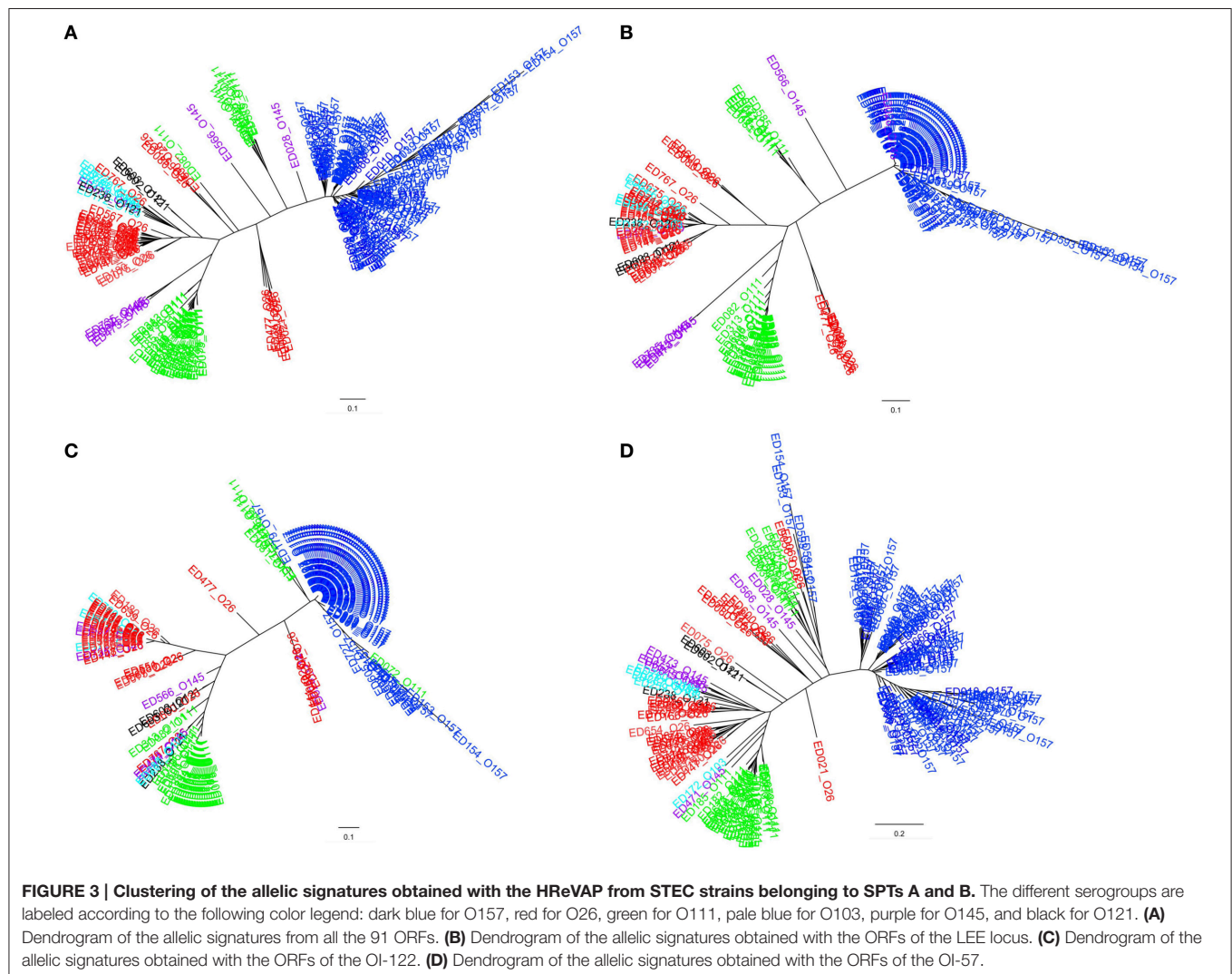


TABLE 1 | List of primers used in this study.

Target pathogenicity island	Gene name	ORF identifier	Primer pair	Localization of the PCR product in the genome of EDL933 strain (Acc. no. AE005174)	Product length (bp)
LEE	rorf1	Z5143	TGCCACCCAGAATAGCCCTGC CATCTGCTTTCCATGCCAGCCC	4691915–4692180	266
	espG	Z5142	AGCTGAAGTTGTGGARTTTTATGC TGTGTCAACRTTAAGGCTGGCA	4691215–4691313	99
	ler	Z5140	CGCCCAGCCAGGTCTGCCC GCAGGAAGCAAAGCGACTGCG	4688879–4689069	191
	orf2	Z5139	AATGAGCAGTTCCTTTGCTTC ACGATAACTGARCTGGAAGAYG	4688596–4688748	153
	cesAB	Z5138	GCTCTTCACCATATGTGTACCCC AAACAGATCCTCAGGCGGCAG	4688331–4688396	66
	orf4	Z5137	CTGGTTTACATTGCCCTCCCT AGATTGCTTGCTCTAAGTAGTGC	4687810–4687960	151
	orf5	Z5136	CATTAAGACCAACTCGATAGCCTGG GAACATTACTGAGGCGGAGGAGG	4687312–4687497	186
	escR	Z5135	TTGCCAGCCTCCAACAAGAATG TCTTATTGGCCTTGGGTATGATG	4686348–4686439	92
	escS	Z5134	CAGTGATAAGTCAGGGCAAGCG GTTCAATTATGTGTGCAAACGTTCTGG	4686113–4686297	185
	escT	Z5133	ACCCAAACCCCATATCACATAAG CGAGATATTACCATTACAAGCC	4685443–4685591	149
	escU	Z5132	ACACTAATCTTCTGAGCCGATGG GAGGTAATGGCTGCAGTGCAGTC	4684910–4685199	290
	rorf3	Z5131	TCTGTTGCTGGTTATATCTTAGCC TCTGAGTTCTACTGCATTGGTG	4684061–4684174	114
	grlR	Z5129	CACAGAAGCAATATCGCCGCC GCAATGAAGACTCCTGTGGGG	4683473–4683549	77
	grlA	Z5128	ATCTCAAGGCGCTTATAATGC GGTATAAACCTGAGGAGAGCA	4682901–4683019	119
	cesD	Z5127	CCAGGATGGTGGCATTAGCG CGCTGATGACATAAGCCCAGAGA	4681955–4682143	189
	escC	Z5126	ACCGGAAAACGTCGGGAAC GTCGCTTCTGTTGTCATTCTG	4681565–4681701	137
	sepD	Z5125	GGATGTCAACGTTACCAACGC CGGCCAGGATATTGCAGAGTCTG	4679969–4680129	161
	escJ	Z5124	AGTGGCATCGCCATCACAAG ACGCTGCAATCGATAACACC	4679610–4679711	102
	rorf8	Z5123	CAGCCCCCTCTGCGGCAGG TGGCGAGAGCCACCTGCAG	4679037–4679206	170
	sepZ	Z5122	TACGGCTCTGGCAGCTCTAGG GCCAGAAGTAATACCCAGGGC	4678718–4678808	91
	orf12	Z5121	GTCATCCTGCGAACGCGCTC CCCTTGCTCCTGAGCGTACG	4678032–4678188	157
	escV	Z5120	GCCAAACAATATTACCGC TTGATGCGCCTGTCGCTCAG	4677681–4677797	117
	escN	Z5119	CGCCCAAGTAGTGCGTCTCC CATGTTTCCGGCATGTACTGTGG	4675680–4675765	86
	orf15	Z5118	TCGCTGCTTTTGCTCATCAAATG GAGTCGATTGCCCTAAGCAG	4674438–4674548	111
	orf16	Z5117	AAACAATTCTCGTAATCTTCCAC CCTCGGGCCGAAGACGACGC	4674052–4674132	81

(Continued)

TABLE 1 | Continued

Target pathogenicity island	Gene name	ORF identifier	Primer pair	Localization of the PCR product in the genome of EDL933 strain (Acc. no. AE005174)	Product length (bp)
	sepQ	Z5116	AGTCTGCATTACCCATAAGATC ACAGTTCATTCACTGGTGCCA	4673626–4673725	100
	cesF	Z5114	GCTCTTGAAGAGCAAACTGACGTG CCAGTATCCATTTCATGGCCATCG	4671935–4671998	64
	tir	Z5112	GATCCCGGCGCTGGTGGG CATGGGAGGATTAACGGGGG	4668929–4668997	69
	cesT	Z5111	GGGGTAGCATCATCGAGAGGG GAATGGTGGGCCATATCTGTGC	4668459–4668542	84
	eae	Z5110	CTCATGCGAAATAGCCGTTA CATTGATCAGGATTTTCTGGTGATA	4667314–4667415	102
	escD	Z5109	AATGGTGCATGCGTAATCGGG GATTGAGGCCTTGTTCAAGGGG	4663987–4664092	106
	sepL	Z5108	CATAAACCAGTCAAGTAAAGAGGC CAGGAAGGTATCCAGAAGATCAA	4663194–4663283	90
	espA	Z5107	ATCAGTGCTACTCTGAACATCAGC GGTAATATGTCGAAGGATGAGGTGG	4662413–4662628	216
	espD	Z5106	TGCAGYGCCATTGCTGTTGC GTAAAATTSTTGGYCAAGTCTTTG	4661487–4661604	118
	espB	Z5105	GWYGTATATCACGCAGACG CGAARACRTTGCCAACAACRRTATCTG	4660090–4660258	169
	cesD2	Z5104	AGCCCTGTTTGTTACGTGC ATCGGATGGCGGATTGAGAC	4659823–4659923	101
	escF	Z5103	CAGGGTCATTAACCAAATCGGTGC CTCAACAAATGGGTGAAGTAGG	4659474–4659557	84
	orf29	Z5102	GCCTTTACCGCATCATTTTTTC CATAAGCTTCCCAAATATGCRGA	4659159–4659289	131
	espF	Z5101	ACGCACGGCCTGAGGGGC TTCTACACTAGGGCGGCAGC	4658870–4658960	91
OI-122		Z4318	TTTTGCCGAATACCTCAGG ATGAGAATGCGGTGGAAG	3924458–3924551	94
		Z4320	ATGTCGGTTTTTCCGTCCA TCACGACTGGAGAGCTTTCTG	3924862–3924936	75
	pagC	Z4321	ATCGATATTGCAGATCACTCC GGCTGATAATCATACGCTATCG	3925569–3925657	89
		Z4322	ATTCCGCACATCTCTGTGGCTA AATGCTGCACTGGCGGTGGT	3926293–3926396	104
		Z4325	TCGACAAAACGCTACAGGAAG CTCAGTGTTCCGTTAAAATGCTC	3928384–3928483	100
		Z4326	TCCTGGATTATTTCTGCATTTC ACTATTGCCAAGTACGCCACAA	3929758–3929833	76
		Z4327	CATCTATTTCTCACGCACTGTAG CAATCAGCAGAGTAATGTTGTC	3930611–3930733	123
	nleB	Z4328	CATGTTGAAGGCTGGAATTTGT CCGCTACAGGGCGATATGTT	3931502–3931573	72
	nleE	Z4329	AGAAGCGTTTGAACCTATTTCCA TTGGGCGTTTTCCGGATAT	3932207–3932289	83

(Continued)

TABLE 1 | Continued

Target pathogenicity island	Gene name	ORF identifier	Primer pair	Localization of the PCR product in the genome of EDL933 strain (Acc. no. AE005174)	Product length (bp)
	efa1	Z4331	TGGAAACCGTGAGCCCATTC TTCTGCGACCTGACCATCCT	3933942–3934035	94
		Z4332	TTTTACCAAGTTCATCATAAGG CCATTATAAACATTTGCCAGACC	3934766–3934848	83
	efa2	Z4333	ACTAAGATCAATACAAGGATTCC ATCCATCAGGCCATAGGTG	3936424–3936496	73
OI-57		Z2036	ATCGCCCGTTTGCTGAGCTT ACGTTTTGGCACCAGACCGT	1850106–1850256	151
		Z2037	TTTCCGGTTCCTGTTGCGCT ACCCGCGATTCTGTGAACCA	1851736–1851871	136
		Z2039	TTTGTGATGCGGTGCCTGGT GGGATTCTCTGTATCCGGCGTT	1853872–1853944	73
		Z2045	ATCGGTTGCGCTGGTTGCTT AGGCCGCATTAGGTAAGCTGGT	1855996–1856164	169
		Z2046	AGCTTGCCAATGTCGAGGA TTCATTGTTCAACCGCCCCG	1856345–1856517	173
		Z2048	TGGCTTTGCCGAGACAGAA TTTAACCTGCGCCCTGACGT	1857700–1857842	143
	adfO	Z2053	AAGTGTGCGCGCAATCGAA GTCTGGCGCTATTTCCACGACA	1860616–1860778	163
		Z2054	AAGGTCAAGGAGAAGCAGGCT TCTTTCCTCGTTACCAAGTCCCGT	1861231–1861334	104
		Z2056	TGACTGGCTGTTGCGTCATGT TGCCAGCACAAACACCATTGC	1862484–1862594	111
		Z2057	TATCAAAAGCCGGGAGCGT TTTTATTGCCAGCCGTCCGGA	1863371–1863465	95
		Z2060	ATGCGGAGCTGCAGAGTGAA TTCTGCCGTTTTTCGCACG	1865516–1865652	137
		Z2065	CACAGCAACCTGCGCTTGTT TGGCACTGCGCGTTAAACAC	1867369–1867445	77
		Z2066	TTACGGTGCGGCATCGAGAA TGCGCGCCCATGAACTGAAA	1867659–1867735	77
		Z2069	TGTCATACACCACGTCAGCGGT ATGTCAGCAGCCCAACAGCA	1868962–1869084	123
		Z2071	TTCTGGCGCTGATTGGTGCA ACGGTATGCTGTGGTGTGGT	1869583–1869674	92
		Z2073	ATGCTCTCAGCCATCGCGTT ACATTCGTGCTGGATGCGCT	1871270–1871454	185
		Z2082	AACGCCATCGAGTCGCTGAA TTGCCAGCCACACGACCTTT	1878129–1878228	100
		Z2084	ATGGTTGGGGCAAACGCTT AACCAGATGGGCCATGCAGA	1878830–1878951	122
		Z2085	TCTTCGCCGTTGACGTTGGT AAAAGGCGGGGAAGATGGCA	1880218–1880319	102
		Z2086	CCGGTTACTGAGCAGAACACCA TTAACGTACCTGGAGGCACCA	1880974–1881105	132
		Z2090	ACAGGCGTTGTGGTCCCTTT TGTGTGCAGCTTGTGCGGAGA	1883045–1883229	185

(Continued)

TABLE 1 | Continued

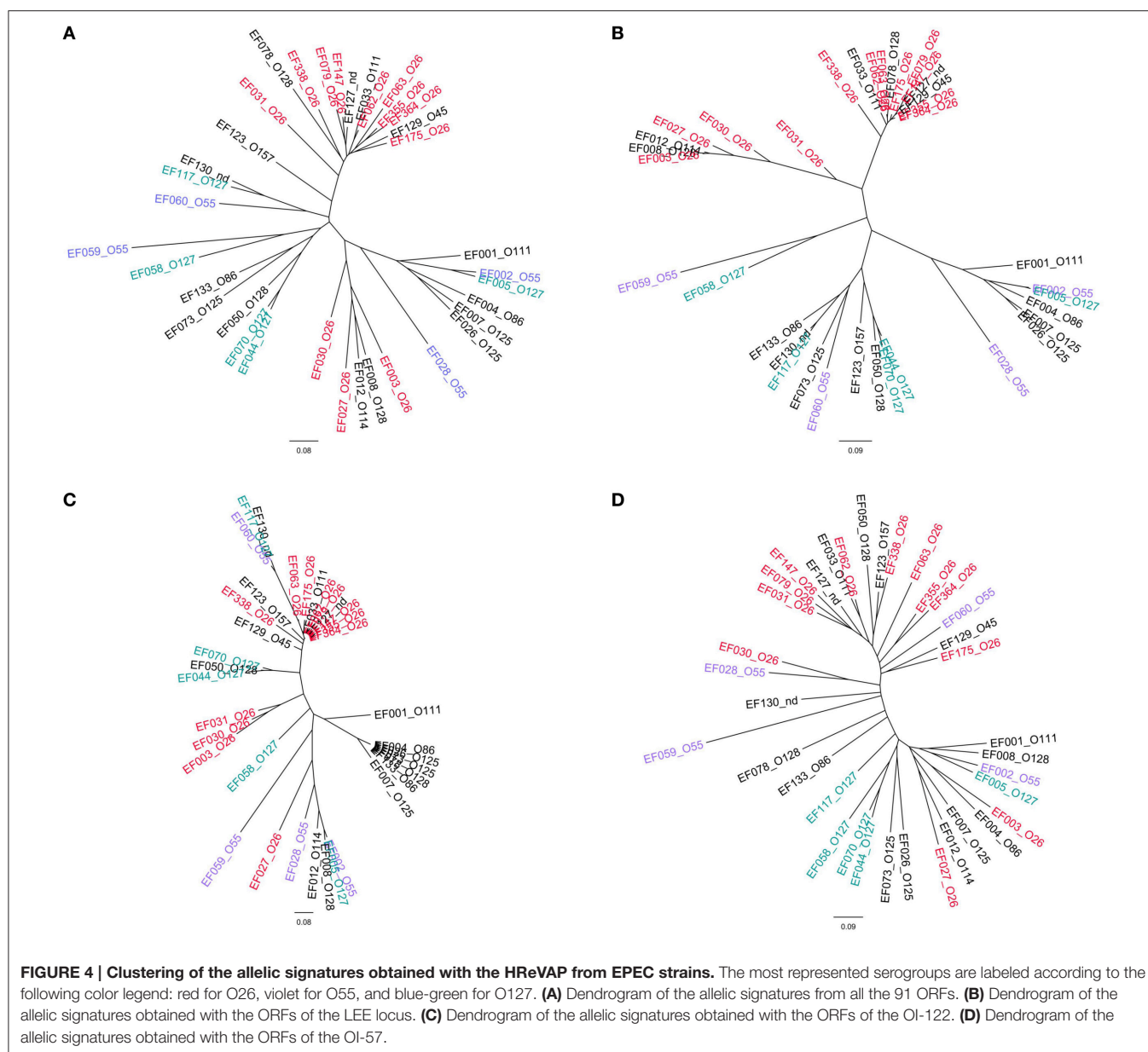
Target pathogenicity island	Gene name	ORF identifier	Primer pair	Localization of the PCR product in the genome of EDL933 strain (Acc. no. AE005174)	Product length (bp)
		Z2091	TGGGCGTTGCTAATTCAGGCA TCGGGAATGACAACCGTCGGTA	1883756–1883832	77
		Z2093	TGGTGAGCGCGTCACGAAAT TCCGTCACATCAAGCGCACT	1884483–1884585	103
		Z2094	TTCAGGAAATCCGGGCTGCA AGGCCGTTTCCGTTGCTGAT	1885233–1885302	70
		Z2096	ACAAAACGGCGGATGAGCTG ACGCCCTTTGCTGATTGCCA	1886138–1886226	89
		Z2100	ATGAGCGGGTGATTCGTGCA ACAACACCATTGCGCCAGCA	1889873–1890019	147
		Z2101	TTGTTTTGCCTTACCCGCCGAC TAAGTGCCACATCCCGGCGATA	1890743–1890847	105
		Z2102	AACGGCAATTAACGAGGCGCT TCGGTACAATGTGCCAGCCA	1891268–1891390	123
		Z2104	TGCGTTGGCTGCGGCTTTAA TCGCCCTGACTTTGGCACTT	1892470–1892627	158
		Z2108	AAAGGGCAGTACGGCTGTGT AATGCCGAACCAACGACACA	1894801–1894916	116
		Z2109	AAAACCTGGTCCCTGACGCA TTGTTGGTCAGTGCGCCTGA	1895736–1895839	104
		Z2112	ACGTTACGGGCCATTTCGA TGTTTGCCATGTTGCGCGGT	1897682–1897770	89
		Z2114	TTTCCCGCCGGCCTTATGT AGACTGCGTTCAAGGCGT	1899877–1899952	76
		Z2116	TTCCTGCGGTTGCGCTCAT ATTTTGATGGAGGGGCCAGCT	1901245–1901344	100
		Z2118	TTTTCAGCTCTCGCATCGGC GCTGGCGAACGCGACAATTA	1903499–1903600	102
		Z2120	AGCGCCTTGTCACGTTGAT ATCTGCCGTGGTGCCATTCT	1904347–1904453	107
		Z2121	GCTGATGCTTCAACGGCTGAA GCTGACATGCGTAACGAGAATCC	1904897–1904991	95
		Z2131	TTGACTCGCAGCGAAACCGT ATCGCCGCATCCACACGTAA	1910373–1910508	136
		Z2146	AAGTGATGACGGTCGCCACA TGCCGGAGCCTTCATAAGCA	1925830–1925936	107
		Z2150	TGCAGGATGCAGCCAGACTTGA TTTCCAGAGCCACAGCCCTT	1928464–1928640	177
		Z2152	GGCCATCGTTGCTGGTGATTT TCTGTTCTGCCCCTGCAACA	1930016–1930109	94

to that produced with the WG-SNP (Figure 6C). Apparently, the topology of the HReVAP dendrogram allowed identifying differences within the serogroups, that were not visible in the WG-SNP-based dendrogram.

DISCUSSION

The detection of the enzyme isoforms or of the polymorphisms in the genomes of pathogenic microorganisms has been for

a long time the basis for the identification of molecular profiles of isolates. Bacterial subtyping has been largely used in research studies on the evolution of bacterial pathogens since the first development of typing methods such as the multi locus enzyme electrophoresis (Selander and Levin, 1980; Selander et al., 1986; Donkor, 2013) and the pulsed field gel electrophoresis (Arbeit et al., 1990), followed by the elaboration of schemes for the identification of the allelic forms of genes as in the multi-locus sequence typing (MLST) (Maiden et al.,



1998). Moreover, molecular typing of microorganisms soon demonstrated its great potential in the control of infectious diseases through the implementation of surveillance programs aiming at limiting the burden of infections (Swaminathan et al., 2001). Nowadays, molecular subtyping of bacteria can benefit from cutting edge technologies such as the next generation sequencing (NGS) allowing the detection of whole genome single nucleotide polymorphism (SNP) (WG-SNP) (Kuroda et al., 2010; Vogler et al., 2011; Joensen et al., 2014; Dallman et al., 2015b).

WG-SNP has been successfully used to define a typing scheme for the surveillance of *Listeria monocytogenes* infections (Commission Decision, 2010). However, its application to bacterial pathogens such as *E. coli*, although advisable, is still under debate given the extensive genomic variability of this

bacterial species. The development of WG-SNP-based typing concepts for *E. coli* has been attempted by several authors and proved successful for some STEC serogroups such as O157 and, to a lesser extent, O26 (Dallman et al., 2015a,b,c; Holmes et al., 2015; Jenkins et al., 2015). However, a single approach successfully applicable to all the STEC serogroups has not been developed yet.

We have deployed a typing scheme for STEC based on the evaluation of polymorphisms in the sequence of a large panel of virulence genes through the determination of the melting temperature of Real-Time PCR amplicons. Such an approach originated from multiple considerations. The virulence genes, part of the accessory genome, have a higher variability than the rest of the genome. Such an increased variability would reduce the number of targets needed for the phylogenetic analysis.

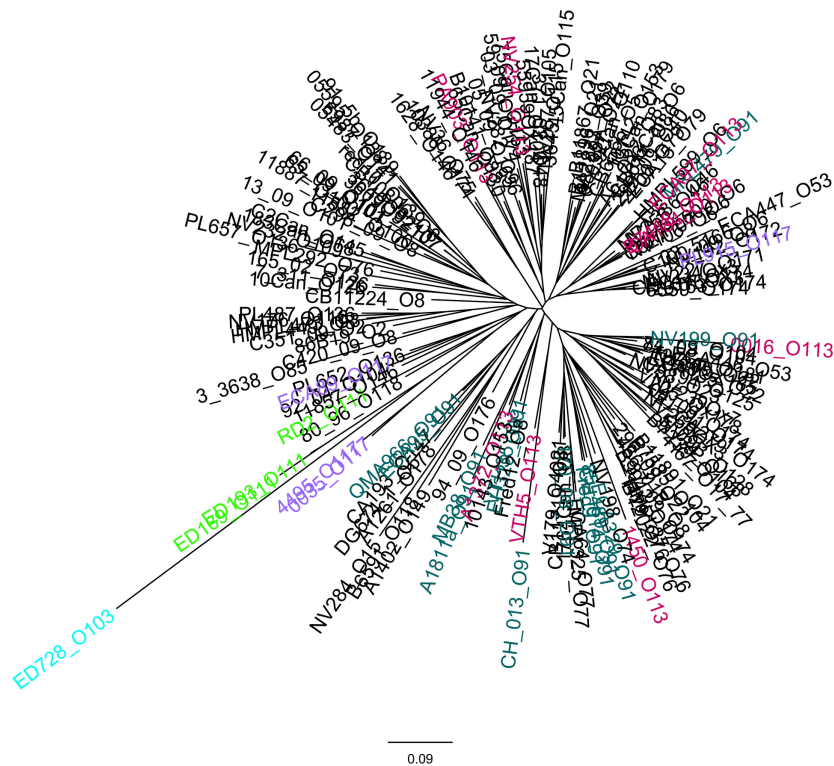


FIGURE 5 | Clustering of the allelic signatures obtained with the HReVAP from the *eae*-negative STEC strains. Dendrogram of the allelic signatures obtained with the ORFs of the OI-57. The most represented serogroups are labeled according to the following color legend: crimson for O113, green for O111, pale blue for O103, orange for O91, and brown for O117.

Additionally, comparing the allelic combinations of the fraction of genome shared by all the members of a pathotype (e.g., STEC) should overcome the need of finding an appropriate reference or setting a threshold for the diversity, which would introduce a bias in the evaluation of clusters. These aspects both represent limitations of the currently described whole genome sequence-based methods for *E. coli* (Dallman et al., 2015a,b,c; Holmes et al., 2015; Jenkins et al., 2015). Finally, the use of the widely diffused Real Time PCR to obtain the strains' signatures makes it not necessary to resort to NGS, which is only available in reference laboratories and requires skills and knowledge of the downstream bioinformatics applications that might be unavailable in most of the front-line laboratories.

The rationale behind the proposed concept resides in the many studies published on the allelic variants of known virulence genes of STEC, such as the Shiga-Toxin-coding genes (Friedrich et al., 2002; Bielaszewska et al., 2006; Persson et al., 2007; Scheut et al., 2012) and the *eae* gene (Oswald et al., 2000; Tarr et al., 2002; Ito et al., 2007; Madic et al., 2010), as well as the more recently described *subAB* and *toxB* genes (Tozzoli et al., 2010; Michelacci et al., 2013, 2014). All the mentioned papers described the association between specific alleles and sub-populations of STEC strains. We investigated the allelic forms of 91 virulence genes conveyed by the three main MGEs associated with STEC pathogenicity, namely the LEE, the OI-122, and the OI-57

(McDaniel and Kaper, 1997; Karmali et al., 2003; Imamovic et al., 2010) and used the obtained allelic signatures to investigate the phylogenesis of STEC.

The whole process, termed High Resolution Virulence Allelic Profiling (HReVAP), allowed us to identify a range of 2–10 allelic forms for each of the 91 ORFs, resulting in the impressive number of 476 total alleles generating a high number of unique allelic signatures.

The HReVAP clustered the LEE-positive STEC strains into groups approximately segregating with the serogroup, providing an indication that the allelic signatures were not randomly assigned to the isolates. Additionally, the analysis identified different subpopulations within the serogroups and also showed variability within each of the populations identified (Figure 3). This finding was not unexpected, since all the strains used in the test panel were epidemiologically unrelated, and at the same time provided an indication that the HReVAP might also be successful in identifying clusters of related strains such as those derived from an outbreak.

The HReVAP produced allelic signatures also with *eae*-negative STEC. However, the dendrogram obtained with these strains had a less resolved topology (Figure 5). An explanation of this result resides either in the lower number of genes these isolates are positive for or in the low number of strains in each serogroup, which in some cases only included one isolate. Nevertheless, the finding that at least part of this PAI

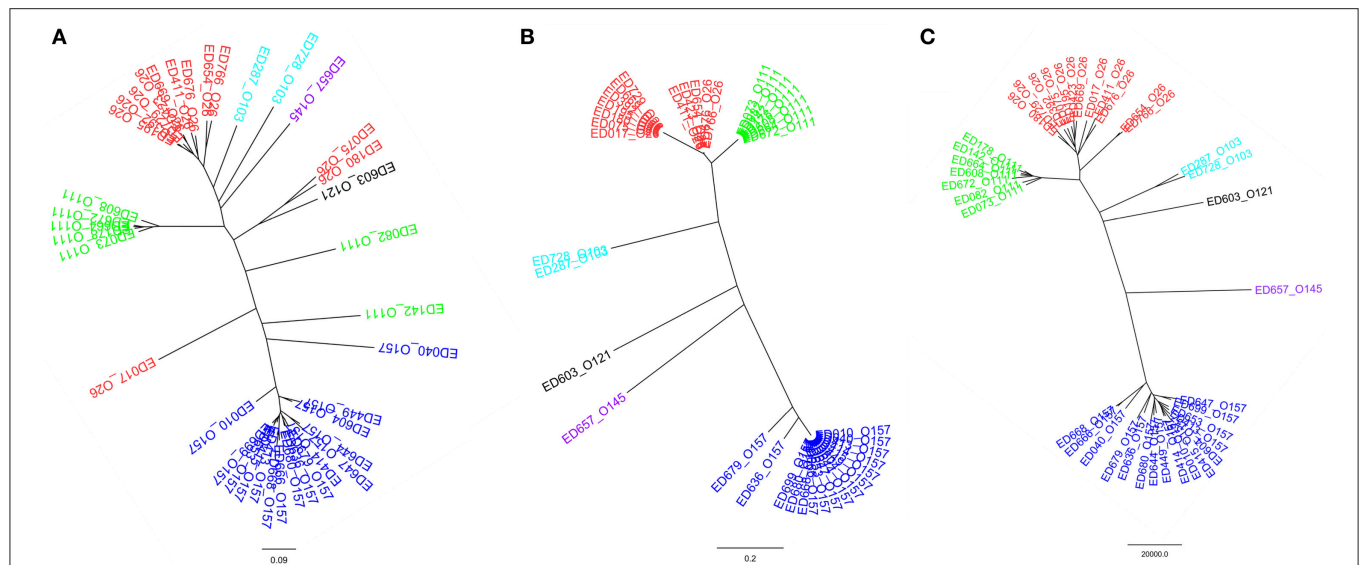


FIGURE 6 | Comparison of the dendrograms obtained by HReVAP, MLST, and WG-SNP typing. The different serogroups are labeled according to the following color legend: dark blue for O157, red for O26, green for O111, pale blue for O103, purple for O145, and black for O121. **(A)** Dendrogram of the allelic signatures obtained with the HReVAP typing. **(B)** Dendrogram obtained from the MLST typing. **(C)** Dendrogram obtained from the WG-SNP typing.

was frequently present in *eae*-negative STEC is interesting and constitutes the first report of the presence of this PAI, or its remnants, in this group of STEC.

The HReVAP analysis also proved useful in following the evolution of the single MGEs considered for the typing scheme. As a matter of fact, we could visualize a similar pattern of variation in the allelic signatures obtained considering the ORFs of the LEE locus and the OI-122 (**Figures 3B,C**). This result indicates that the two MGEs underwent similar evolutionary pathways and supports the previous hypothesis about their common acquisition through a single event of horizontal gene transfer in certain STEC and EPEC strains (Morabito et al., 2003).

Our results showed that the OI-57 had the greatest genetic variability, displaying the highest number of alleles on average for all the ORFs considered (**Figure 1C** and Supplementary Table 3). Additionally, the analysis of the allelic signatures obtained considering the OI-57 ORFs produced dendrograms with the most dispersed topology (**Figures 3D, 4D**). These observations suggest that this MGE could have been acquired at an early stage of the evolutionary pathway that led to the emergence of STEC. Additionally, since the LEE-negative STEC investigated were also positive for many of the OI-57-related ORFs considered in this study, it can be hypothesized that such an island could be a common heritage of STEC independently of the presence of the LEE locus.

Finally, the comparison of the performance of the HReVAP with that obtained with other comparative genomic tools such as the MLST and the WG-SNP analysis substantiates the robustness of the HReVAP in identifying LEE-positive STEC populations with a much higher resolution with respect to the MLST and a comparable level of discrimination to that of the WG-SNP.

In conclusion, the HReVAP approach demonstrated good sensitivity and high resolution in the molecular characterization of STEC, particularly for the LEE-positive strains. Moreover, the incredibly large virulome of pathogenic *E. coli* offers the opportunity to refine the HReVAP typing strategy for other STEC groups, such as the LEE-negative isolates, or even to extend it to other *E. coli* pathotypes by integrating the panel of targets.

Further work is in progress to assess the use of HReVAP as an effective tool for the surveillance of STEC infections and to obtain the allelic signatures from whole genome sequences in order to make this technique a cross-generational tool connecting the Real-Time PCR and the NGS-based applications.

AUTHOR CONTRIBUTIONS

VM conceived the experimental design and drafted the manuscript, MO developed the scripts for HReVAP clustering, and critically revised the manuscript, AK developed and applied the scripts for the extraction of the allelic signatures of the HReVAP and critically revised the manuscript, SD and PF designed the Real Time PCR primers, performed the amplifications and melting curve analyses and participated in the revision of the manuscript, AC contributed to the revision of the draft manuscript for important intellectual content, SM conceived the study and thoroughly revised the manuscript. Finally, all the authors approved the manuscript to be published.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00202>

REFERENCES

- Arbeit, R. D., Arthur, M., Dunn, R., Kim, C., Selander, R. K., and Goldstein, R. (1990). Resolution of recent evolutionary divergence among *Escherichia coli* from related lineages: the application of pulsed field electrophoresis to molecular epidemiology. *J. Infect. Dis.* 161, 230–235. doi: 10.1093/infdis/161.2.230
- Bielaszewska, M., Friedrich, A. W., Aldick, T., Schurk-Bulgrin, R., and Karch, H. (2006). Shiga toxin activatable by intestinal mucus in *Escherichia coli* isolated from humans: predictor for a severe clinical outcome. *Clin. Infect. Dis.* 43, 1160–1167. doi: 10.1086/508195
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19:Unit 19.10.11–21. doi: 10.1002/0471142727.mb1910s89
- Caprioli, A., Morabito, S., Brugere, H., and Oswald, E. (2005). Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet. Res.* 36, 289–311. doi: 10.1051/vetres:2005002
- Commission Decision, E. C. (2010). Commission decision concerning a financial contribution from the Union towards a coordinated monitoring programme on the prevalence of *Listeria monocytogenes* in certain ready-to-eat foods to be carried out in the Member States (2010/678/EU). *Official J. Eur. Union, L* 292, 40–54. doi: 10.3000/17252555.L_2010.292.eng
- Dallman, T. J., Byrne, L., Ashton, P. M., Cowley, L. A., Perry, N. T., Adak, G., et al. (2015a). Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin. Infect. Dis.* 61, 305–312. doi: 10.1093/cid/civ318
- Dallman, T. J., Byrne, L., Launders, N., Glen, K., Grant, K. A., and Jenkins, C. (2015b). The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. *Epidemiol. Infect.* 143, 1672–1680. doi: 10.1017/S0950268814002696
- Dallman, T. J., Ashton, C., Byrne, L., Perry, N. T., Petrovska, L., Ellis, R., et al. (2015c). Applying phylogenomics to understand the emergence of Shiga-toxin-producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK. *Microb. Genomics* 1. doi: 10.1099/mgen.0.000029
- Donkor, E. S. (2013). Molecular typing of the pneumococcus and its application in epidemiology in sub-Saharan Africa. *Front. Cell. Infect. Microbiol.* 3:12. doi: 10.3389/fcimb.2013.00012
- Drummond, A., Heled, J., Lemey, P., de Oliveira, T., Pybus, O., Shapiro, B., et al. (2012). *FigTree Graphical Viewer of Phylogenetic Trees* [Online]. Available online at: <http://tree.bio.ed.ac.uk/software/figtree/>
- Friedrich, A. W., Bielaszewska, M., Zhang, W. L., Pulz, M., Kuczius, T., Ammon, A., et al. (2002). *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J. Infect. Dis.* 185, 74–84. doi: 10.1086/338115
- 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
- Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455. doi: 10.1101/gr.4086505
- Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86
- Griffin, P. M., and Tauxe, R. V. (1991). The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiol. Rev.* 13, 60–98.
- Hernandez, D., Francois, P., Farinelli, L., Osteras, M., and Schrenzel, J. (2008). *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.* 18, 802–809. doi: 10.1101/gr.072033.107
- Ho, T. D., Davis, B. M., Ritchie, J. M., and Waldor, M. K. (2008). Type 2 secretion promotes enterohemorrhagic *Escherichia coli* adherence and intestinal colonization. *Infect. Immun.* 76, 1858–1865. doi: 10.1128/IAI.01688-07
- Holmes, A., Allison, L., Ward, M., Dallman, T. J., Clark, R., Fawkes, A., et al. (2015). Utility of whole-genome sequencing of *Escherichia coli* O157 for outbreak detection and epidemiological surveillance. *J. Clin. Microbiol.* 53, 3565–3573. doi: 10.1128/JCM.01066-15
- Imamovic, L., Tozzoli, R., Michelacci, V., Minelli, F., Marziano, M. L., Caprioli, A., et al. (2010). O157, a genomic island of *Escherichia coli* O157, is present in other seropathotypes of Shiga toxin-producing *E. coli* associated with severe human disease. *Infect. Immun.* 78, 4697–4704. doi: 10.1128/IAI.00512-10
- Ito, K., Iida, M., Yamazaki, M., Moriya, K., Moroishi, S., Yatsuyanagi, J., et al. (2007). Intimin types determined by heteroduplex mobility assay of intimin gene (eae)-positive *Escherichia coli* strains. *J. Clin. Microbiol.* 45, 1038–1041. doi: 10.1128/JCM.01103-06
- Jenkins, C., Dallman, T. J., Launders, N., Willis, C., Byrne, L., Jorgensen, F., et al. (2015). Public health investigation of two outbreaks of shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. *Appl. Environ. Microbiol.* 81, 3946–3952. doi: 10.1128/AEM.04188-14
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Karmali, M. A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., et al. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *J. Clin. Microbiol.* 41, 4930–4940. doi: 10.1128/JCM.41.11.4930-4940.2003
- Kuroda, M., Serizawa, M., Okutani, A., Sekizuka, T., Banno, S., and Inoue, S. (2010). Genome-wide single nucleotide polymorphism typing method for identification of *Bacillus anthracis* species and strains among *B. cereus* group species. *J. Clin. Microbiol.* 48, 2821–2829. doi: 10.1128/JCM.00137-10
- Madic, J., Peytavin de Garam, C., Vingadassalon, N., Oswald, E., Fach, P., Jamet, E., et al. (2010). Simplex and multiplex real-time PCR assays for the detection of flagellar (H-antigen) fliC alleles and intimin (eae) variants associated with enterohaemorrhagic *Escherichia coli* (EHEC) serotypes O26:H11, O103:H2, O111:H8, O145:H28 and O157:H7. *J. Appl. Microbiol.* 109, 1696–1705. doi: 10.1111/j.1365-2672.2010.04798.x
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140
- McDaniel, T. K., and Kaper, J. B. (1997). A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol. Microbiol.* 23, 399–407. doi: 10.1046/j.1365-2958.1997.2311591.x
- Michelacci, V., Grande, L., Tozzoli, R., Maugliani, A., Caprioli, A., and Morabito, S. (2014). Identification of two allelic variants of toxB gene and investigation of their distribution among Verocytotoxin-producing *Escherichia coli*. *Int. J. Med. Microbiol.* 304, 730–734. doi: 10.1016/j.ijmm.2014.05.009
- Michelacci, V., Tozzoli, R., Caprioli, A., Martinez, R., Scheutz, F., Grande, L., et al. (2013). A new pathogenicity island carrying an allelic variant of the Subtilase cytotoxin is common among Shiga toxin producing *Escherichia coli* of human and ovine origin. *Clin. Microbiol. Infect.* 19, E149–E156. doi: 10.1111/1469-0691.12122
- Morabito, S., Tozzoli, R., Oswald, E., and Caprioli, A. (2003). A mosaic pathogenicity island made up of the locus of enterocyte effacement and a pathogenicity island of *Escherichia coli* O157:H7 is frequently present in attaching and effacing *E. coli*. *Infect. Immun.* 71, 3343–3348. doi: 10.1128/IAI.71.6.3343-3348.2003
- Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.
- O'Brien, A. D., and Holmes, R. K. (1987). Shiga and Shiga-like toxins. *Microbiol. Rev.* 51, 206–220.
- O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W., and Formal, S. B. (1984). Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* 226, 694–696. doi: 10.1126/science.6387911
- Oswald, E., Schmidt, H., Morabito, S., Karch, H., Marches, O., and Caprioli, A. (2000). Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. *Infect. Immun.* 68, 64–71. doi: 10.1128/IAI.68.1.64-71.2000

- Paton, A. W., Srimanote, P., Woodrow, M. C., and Paton, J. C. (2001). Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. *Infect. Immun.* 69, 6999–7009. doi: 10.1128/IAI.69.11.6999-7009.2001
- Perelle, S., Dilasser, F., Grout, J., and Fach, P. (2004). Detection by 5'-nuclease PCR of Shiga-toxin producing *Escherichia coli* O26, O55, O91, O103, O111, O113, O145 and O157:H7, associated with the world's most frequent clinical cases. *Mol. Cell. Probes* 18, 185–192. doi: 10.1016/j.mcp.2003.12.004
- Persson, S., Olsen, K. E., Ethelberg, S., and Scheutz, F. (2007). Subtyping method for *Escherichia coli* shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J. Clin. Microbiol.* 45, 2020–2024. doi: 10.1128/JCM.02591-06
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- RTeam, C. (2014). *R: A Language and Environment for Statistical Computing* [Online]. Available online at: <http://www.R-project.org/>
- Scheutz, F., Teel, L. D., Beutin, L., Pierard, D., Buvens, G., Karch, H., et al. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12
- Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N., and Whittam, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* 51, 873–884.
- Selander, R. K., and Levin, B. R. (1980). Genetic diversity and structure in *Escherichia coli* populations. *Science* 210, 545–547. doi: 10.1126/science.6999623
- Stover, B. C., and Muller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11:7. doi: 10.1186/1471-2105-11-7
- Swaminathan, B., Barrett, T. J., Hunter, S. B., Tauxe, R. V., and CDC PulseNet Task Force (2001). PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg. Infect. Dis.* 7, 382–389. doi: 10.3201/eid0703.017303
- Tarr, C. L., Large, T. M., Moeller, C. L., Lacher, D. W., Tarr, P. I., Acheson, D. W., et al. (2002). Molecular characterization of a serotype O121:H19 clone, a distinct Shiga toxin-producing clone of pathogenic *Escherichia coli*. *Infect. Immun.* 70, 6853–6859. doi: 10.1128/IAI.70.12.6853-6859.2002
- Tozzoli, R., Caprioli, A., Cappannella, S., Michelacci, V., Marziano, M. L., and Morabito, S. (2010). Production of the subtilase AB5 cytotoxin by Shiga toxin-negative *Escherichia coli*. *J. Clin. Microbiol.* 48, 178–183. doi: 10.1128/JCM.01648-09
- Vogler, A. J., Chan, F., Wagner, D. M., Roumagnac, P., Lee, J., Nera, R., et al. (2011). Phylogeography and molecular epidemiology of *Yersinia pestis* in Madagascar. *PLoS Negl. Trop. Dis.* 5:e1319. doi: 10.1371/journal.pntd.0001319
- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., et al. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60, 1136–1151. doi: 10.1111/j.1365-2958.2006.05172.x

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

Copyright © 2016 Michelacci, Orsini, Knijn, Delannoy, Fach, Caprioli and Morabito. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Revisiting the STEC Testing Approach: Using *espK* and *espV* to Make Enterohemorrhagic *Escherichia coli* (EHEC) Detection More Reliable in Beef

Sabine Delannoy¹, Byron D. Chaves², Sarah A. Ison², Hattie E. Webb², Lothar Beutin³, José Delaval⁴, Isabelle Billet⁵ and Patrick Fach^{1*}

¹ Food Safety Laboratory, Université Paris-Est, Anses (French Agency for Food, Environmental and Occupational Health and Safety), Platform IdentityPath, Maisons-Alfort, France, ² Department of Animal and Food Sciences, Texas Tech University, Lubbock, TX, USA, ³ Division of Microbial Toxins, National Reference Laboratory for *Escherichia coli*, Federal Institute for Risk Assessment, Berlin, Germany, ⁴ Laboratoire de Touraine, (LDA37) Conseil Départemental, Tours, France, ⁵ SAS Charal Groupe Bigard, Cholet, France

OPEN ACCESS

Edited by:

Pina Frattamico,
United States Department of
Agriculture, Agricultural Research
Service, USA

Reviewed by:

James L. Smith,
United States Department of
Agriculture, USA
George Carl Paoli,
United States Department of
Agriculture, USA

*Correspondence:

Patrick Fach
patrick.fach@anses.fr

Specialty section:

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

Received: 30 October 2015

Accepted: 05 January 2016

Published: 22 January 2016

Citation:

Delannoy S, Chaves BD, Ison SA,
Webb HE, Beutin L, Delaval J, Billet I
and Fach P (2016) Revisiting the
STEC Testing Approach: Using *espK*
and *espV* to Make Enterohemorrhagic
Escherichia coli (EHEC) Detection
More Reliable in Beef.
Front. Microbiol. 7:1.
doi: 10.3389/fmicb.2016.00001

Current methods for screening Enterohemorrhagic *Escherichia coli* (EHEC) O157 and non-O157 in beef enrichments typically rely on the molecular detection of *stx*, *eae*, and serogroup-specific *wzx* or *wzy* gene fragments. As these genetic markers can also be found in some non-EHEC strains, a number of “false positive” results are obtained. Here, we explore the suitability of five novel molecular markers, *espK*, *espV*, *ureD*, *Z2098*, and CRISPR_{O26:H11} as candidates for a more accurate screening of EHEC strains of greater clinical significance in industrialized countries. Of the 1739 beef enrichments tested, 180 were positive for both *stx* and *eae* genes. Ninety (50%) of these tested negative for *espK*, *espV*, *ureD*, and *Z2098*, but 12 out of these negative samples were positive for the CRISPR_{O26:H11} gene marker specific for a newly emerging virulent EHEC O26:H11 French clone. We show that screening for *stx*, *eae*, *espK*, and *espV*, in association with the CRISPR_{O26:H11} marker is a better approach to narrow down the EHEC screening step in beef enrichments. The number of potentially positive samples was reduced by 48.88% by means of this alternative strategy compared to the European and American reference methods, thus substantially improving the discriminatory power of EHEC screening systems. This approach is in line with the EFSA (European Food Safety Authority) opinion on pathogenic STEC published in 2013.

Keywords: STEC, O157, non-O157, *espK*, *espV*, *ureD*, *Z2098*, CRISPR

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are important zoonotic pathogens comprising more than 400 serotypes (Beutin and Fach, 2014). A fraction of these serotypes are able to cause bloody diarrhea and may progress to hemolytic uremic syndrome (HUS). This subset of STECs is termed Enterohemorrhagic *E. coli* (EHEC) (Beutin and Fach, 2014). STEC O157:H7 has been the first pathogenic *E. coli* whose presence in foodstuffs was regulated. Today, non-O157 STEC

infections have increased greatly, sometimes accounting for up to 70% of notified STEC infections (Brooks et al., 2005; Johnson et al., 2006; Gould et al., 2013; EFSA, 2014). Consequently, regulations in the US and in the EU have evolved to include some non-O157 STEC serogroups together with serotype O157:H7.

Successful implementation of these regulations in the food industry across the world requires effective detection methods that are both specific and sensitive. Detection of non-O157 STEC in foods is particularly challenging because these bacteria lack phenotypic characteristics that distinguish them invariably from the large number of non-STEC flora that share the same habitat. Additionally, they may be present in very low numbers, with a heterogeneous distribution in the food matrices. The environment of the food matrix may also trigger stress responses from the STEC strains and induce latent physiological states that further complicate detection (Wang et al., 2013).

In the absence of a clear definition of virulent STEC strains, the ISO/CEN 13136:2012 Technical Specification (ISO, 2012) and US MLG5B.05 (USDA-FSIS, 2014) use a stepwise approach, comprising an initial screening step for virulence genes (Shiga toxin genes, *stx*, and intimin gene, *eae*), followed by testing of O-serogroup specific gene markers. Because the *stx* and *eae* genes can be independently present in a number of non-pathogenic strains of *E. coli* and other *Enterobacteriaceae*, the first screening step generates numerous signals from samples that do not necessarily contain a true EHEC strain. These *stx/eae* positive enrichments must then be subjected to a second screening targeting the O-group gene markers. As testing is completed on enrichment broths that contain a mixture of different cells, the different target gene signals (i.e., *stx*, *eae*, and the O-group markers) may arise from different individual non-pathogenic strains. Lastly it is necessary to perform an isolation step to confirm the presence of the different markers in a single isolate by PCR.

We have previously shown that combining the detection of *espK* with either *espV*, *ureD*, or *Z2098* is a highly sensitive and specific approach for identifying the top seven clinically important EHEC serotypes in industrialized countries (Delannoy et al., 2013a). These markers were shown to be preferentially associated with *E. coli* strains carrying *stx* and *eae* genes, known as typical EHEC, and could be used in conjunction with *stx/eae* screening to better identify samples that may be more likely to contain a true EHEC.

Recently, a new clone of STEC O26:H11 harboring *stx2a* and strongly associated with HUS has emerged (Bielaszewska et al., 2013). Characterization of O26:H11 *stx2* circulating in France (Delannoy et al., 2015a,b) demonstrated that some *stx2a* or *stx2d* positive strains do not have any of the *espK*, *espV*, *ureD*, or *Z2098* markers. Hence, such clones would evade a first detection step solely based on *stx/eae* and a combination of those genetic markers. Therefore, we identified a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) sequence specific for this O26:H11 EHEC French clone (Delannoy et al., 2015a). Detection of this CRISPR sequence would advantageously be combined with *stx* and *eae* in the first screening step to identify this new O26:H11 clone. Furthermore, we have previously developed CRISPR-based real-time PCR assays able to detect the top

seven EHEC serotypes and the German O104:H4 STEC clone responsible for a very large European STEC outbreak in 2011 (Delannoy et al., 2012a,b; Miko et al., 2013). These CRISPR PCR assays proved highly sensitive and specific when tested on a large collection of *E. coli* strains comprising various *E. coli* pathogroups (Delannoy et al., 2012a,b). Such CRISPR markers may substitute O group testing in a more targeted second step.

Different approaches have been tested to refine the detection systems for EHEC. Some involve detection of *ecf1*, a plasmid gene highly associated with *E. coli* strains that are positive for *stx*, *eae*, and *ehxA* (Livezey et al., 2015); while others involve detection of specific *eae*-subtypes (Bugarel et al., 2010). The top seven EHEC serotypes are exclusively associated with certain *eae*-subtypes (Oswald et al., 2000; Bugarel et al., 2010; Madic et al., 2010). Indeed, O157:H7 and O145:H28 serotypes are associated with the *eae*-gamma subtype; O26:H11 are associated with the *eae*-beta subtype; O103:H2, O121:H19, and O45:H2 are associated with the *eae*-epsilon subtype; and O111:H8 are associated with the *eae*-theta subtype. Real-time PCR targeting these *eae*-subtypes were previously developed and tested in raw milk cheese (Madic et al., 2011) and cattle feces (Bibbal et al., 2014). These could also be used as targets in a more targeted second step.

We attempted to develop an alternative real-time PCR-based approach to improve the detection of the clinically important EHEC by reducing the number of potential positive samples that require further confirmation of the O-antigen markers. We also aimed at reducing the number of samples for which isolation is attempted, as the isolation step is laborious, time-consuming, and not always successful (Wang et al., 2013). The objective of this project was to evaluate the discriminatory power of these various genetic markers to predict the presence of the top seven EHEC serogroups compared to that of the strategy proposed by the ISO/CEN TS13136:2012 (ISO, 2012) and MLG5B.05 (USDA-FSIS, 2014).

MATERIALS AND METHODS

E. coli Control Strains

E. coli control strains used for this study (Table 1) comprised a panel of ATCC *E. coli* strains ($n = 42$) and *E. coli* reference strains derived from the BfR and Anses collections ($n = 13$). The origin and characteristics of the *E. coli* strains from BfR and Anses have been previously described (Delannoy et al., 2012a, 2013a). Cultivation of bacteria and preparation of DNA was performed as previously described (Delannoy et al., 2012a, 2013a).

Beef Samples, Enrichment, and DNA Extraction

A set of 1739 beef samples composed of ground beef and carcasses were collected from routine screening using the GeneDisc array (Pall GeneDisc, Bruz, France) at the Veterinary Departmental Laboratory of Touraine, France during a 1-year period as well as in meat production plants. For this study, sampling was biased to get greater numbers of DNA samples positive for *stx* alone ($n = 306$), positive for *stx* and *eae* ($n = 180$), positive for *eae* alone ($n = 200$), and negative for both *stx* and *eae* ($n =$

TABLE 1 | PCR-screening of the genetic markers in STEC isolates used as reference.

Strain	Code	Serotype	stx1	stx2	eae	eae-beta	eae-gamma	eae-epsilon	eae-theta	espK	espV	Z2098	ureD	CRISPR _{O26:H11} SP_Q26-E
ATCC-2129	TW07865	O145:H28	–	+	+	–	+	–	–	+	+	+	+	–
ATCC-2192	99-3311	O145	+	+	+	–	+	–	–	+	+	+	+	–
ATCC-2194	2001-3022	O145	–	+	+	–	+	–	–	+	+	+	+	–
ATCC-2195	2002-3034	O145	+	–	+	–	+	–	–	+	+	+	+	–
ATCC-2206	2000-3413	O145	–	+	+	–	+	–	–	+	+	+	+	–
ATCC-2208	2003-3054	O145	–	+	+	–	+	–	–	+	+	+	+	–
ATCC-2181	00-3412	O26:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2183	2011-3139	O26:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2186	99-3294	O26:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2188	99-3301	O26:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2196	2003-3014	O26:H11	+	+	+	+	–	–	–	+	–	+	+	–
ATCC-2204	2001-3234	O26:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2205	2003-3023	O26:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2442	ATCC-2442	O26	+	–	+	+	–	–	–	+	–	+	+	–
CB14699	CB14699	O26:H11	–	+	+	+	–	–	–	–	–	–	–	+
ATCC-2187	99-3300	O121:H19	–	+	+	–	–	+	–	+	+	+	+	–
ATCC-2203	2000-3370	O121:H19	–	+	+	–	–	+	–	+	+	+	+	–
ATCC-2219	2002-3211	O121:H19	–	+	+	–	–	+	–	+	+	+	+	–
ATCC-2220	10C-3041	O121:H19	–	+	+	–	–	+	–	+	+	+	+	–
ATCC-2221	09C-3857	O121:H19	+	+	+	–	–	+	–	+	+	+	+	–
ATCC-43890	CDC-C984	O157:H7	+	–	+	–	+	–	–	–	+	+	+	–
ATCC-43895	EDL933	O157:H7	+	+	+	–	+	–	–	+	+	+	+	–
ATCC-179	CDC1997-3215	O111:H8	+	+	+	–	–	–	+	+	+	+	+	–
ATCC-181	CDC1999-3249	O111:H8	+	+	+	–	–	–	+	+	+	+	+	–
ATCC-184	CDC2000-3025	O111:H8	+	–	+	–	–	–	+	+	+	+	+	–
ATCC-2180	00-3237	O111:H8	+	+	+	–	–	–	+	+	+	+	+	–
ATCC-2201	2002-3092	O111:H8	+	–	+	–	–	–	+	+	+	+	+	–
ATCC-2182	2001-3010	O111	+	–	+	–	–	–	+	+	+	+	+	–
ATCC-2209	2001-3357	O111	+	+	+	–	–	–	+	+	+	+	+	–
ATCC-2440	ATCC-2440	O111	+	+	+	–	–	–	+	+	+	+	+	–
ATCC-2441	ATCC-2441	O111	+	–	+	–	–	–	+	+	+	+	+	–
ATCC-2202	99-3075	O45:H2	+	–	+	–	–	+	–	+	–	+	+	–
ATCC-2185	99-3291	O45:H2	+	–	+	–	–	+	–	+	–	+	+	–
ATCC-2189	99-3303	O45:H2	+	–	+	–	–	+	–	+	–	+	+	–
ATCC-2191	98-3167	O45:H2	+	–	+	–	–	+	–	+	–	+	+	–
ATCC-2193	2000-3039	O45:H2	+	–	+	–	–	+	–	+	–	+	+	–
ATCC-2198	98-3215	O45:H2	+	–	+	–	–	+	–	+	–	+	+	–
ATCC-2200	2001-3225	O103:H11	+	–	+	+	–	–	–	+	+	+	+	–
ATCC-2215	2006-3008	O103:H11	+	–	+	–	–	+	–	+	+	+	+	–
ATCC-2207	2001-3304	O103:H2	+	–	+	–	–	+	–	+	+	+	+	–
ATCC-2210	2003-3112	O103:H2	+	–	+	–	–	+	–	+	+	+	+	–
PMK5	PMK5	O103:H2	+	–	+	–	–	+	–	+	+	+	–	–
CB12062	CB12062	O103:H2	+	–	+	–	–	+	–	+	+	+	–	–
CB12092	CB12092	O103:H2	+	–	+	–	–	+	–	+	+	+	–	–
CB11097	CB11097	O103:H25	–	+	+	–	–	–	+	+	+	+	+	–
ATCC-2213	2005-3546	O103:H25	+	–	+	–	–	–	+	+	+	+	+	–
ATCC-2199	2000-3281	O103:H25	+	–	+	–	–	–	+	+	+	+	+	–
CB11784	ECA34	O5	+	–	+	+	–	–	–	+	–	+	+	–
CB12339	CB12339	O5	+	–	+	+	–	–	–	+	–	+	+	–
CB13683	5906	O55:H7	–	+	+	–	+	–	–	+	+	–	–	–

(Continued)

TABLE 1 | Continued

Strain	Code	Serotype	<i>stx1</i>	<i>stx2</i>	<i>eae</i>	<i>eae</i> -beta	<i>eae</i> -gamma	<i>eae</i> -epsilon	<i>eae</i> -theta	<i>espK</i>	<i>espV</i>	Z2098	<i>ureD</i>	CRISPR _{O26:H11} SP _{O26-E}
7123A	7123A	O55:H7	–	+	+	–	+	–	–	+	+	–	–	–
CB7035	CB7035	O118:H16	+	+	+	+	–	–	–	+	–	+	+	–
CB8255	CB8255	O118:H16	+	–	+	+	–	–	–	+	–	+	+	–
CB9915	CB9915	O123:H11	–	+	+	+	–	–	–	+	–	–	+	–
CB10528	CB10528	O172:H25	–	+	+	–	–	–	–	+	+	–	–	–

1053). This sampling scheme does not represent the prevalence of STEC and EHEC in French beef. All samples were incubated in buffered peptone water (BioMerieux, Marcy l'étoile, France) for 18–24 h at 37°C. After enrichment, DNA was extracted from 1 ml of enriched sample using the InstaGene matrix (Bio-Rad Laboratories, Marnes-La Coquette, France) following manufacturer's instruction and DNA was stored at –20°C until use. When samples were found positive for *stx*, *eae*, and *rfbE*_{O157}, isolation of strains was attempted by local laboratories for confirmation of EHEC O157:H7. Following the recommendation of the French ministry of agriculture the appropriate sanitary measures were taken in positive cases of EHEC O157:H7. Unfortunately, because this study was performed several months after the samples were collected, the original samples were not conserved to attempt isolation from presumptive positives with the alternate methods described in this study.

High-Throughput Real-Time PCR

A LightCycler[®] 1536 (Roche, Meylan, France) was used to perform high throughput real-time PCR amplifications as described previously (Delannoy et al., 2012a), except that 1 µl of sample DNA was used in each reaction for a final reaction volume of 2 µl. The thermal profile was modified as follows: 95°C for 1 min, followed by 45 cycles of 95°C for 0 s, and 58°C for 30 s. All ramp rates were set to 2°C/s. *E. coli* gene targets used for the real-time PCR amplification and all primers and probes that have previously been described are reported in Table 2. An inhibition control (IC) was performed on each sample to check for potential inhibition of the PCR reaction due to intrinsic characteristics of the sample. The IC is a recombinant pBluescript IISK+ plasmid containing the *dsb* gene from *Ehrlichia canis* (Michelet et al., 2014). The plasmid was added to each sample at a concentration of approximately 0.3 pg/µl. Primers and probe specific for the *E. canis dsb* gene were used to detect the IC (Michelet et al., 2014).

RESULTS

Presence of *stx1*, *stx2*, *eae*, *espK*, *espV*, Z2098, and *ureD* in *E. coli* Strains

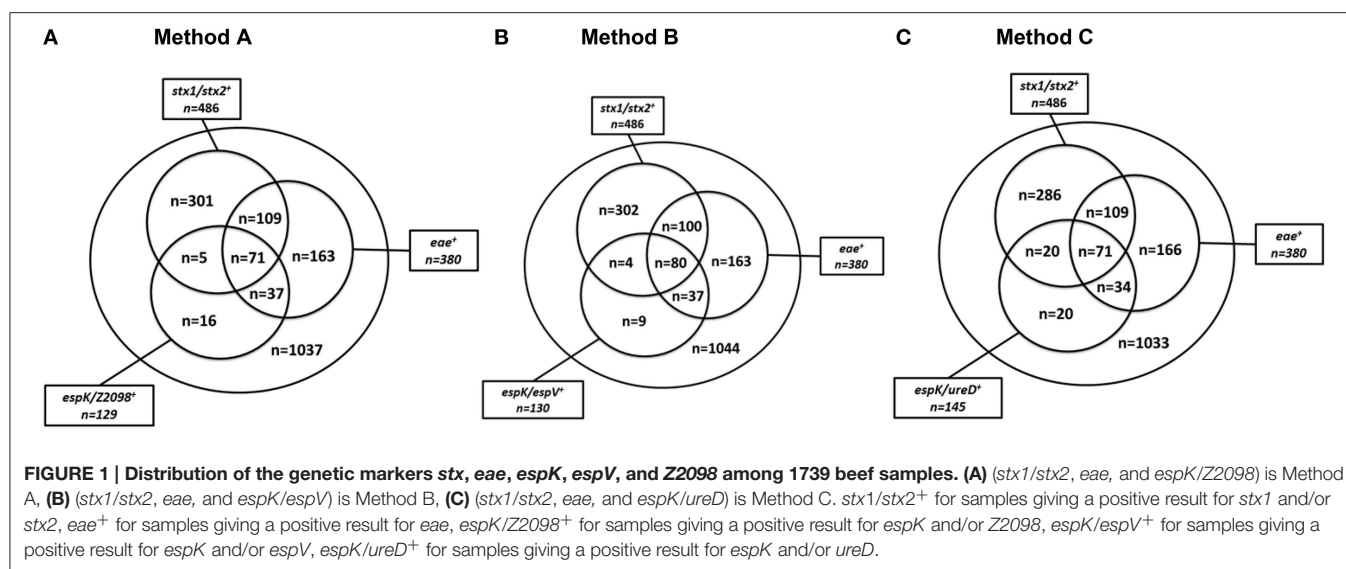
The presence of *stx1*, *stx2*, *eae*, *espK*, *espV*, Z2098, and *ureD* was tested by PCR in a panel of *E. coli* strains obtained from culture collections (Table 1). All *E. coli* strains used in this study were positive for the *stx* and *eae* genes and therefore can be considered as typical EHEC strains. Strains were associated with the following *eae* variants: *eae*-beta (O26:H11, O26:H_{ND},

O103:H11, O5, O118:H16, O123:H11), *eae*-gamma (O145:H28, O145:H_{ND}, O157:H7, O55:H7), *eae*-epsilon (O121:H19, O45:H2, O103:H2, O103:H11), and *eae*-theta (O111:H8, O111:H_{ND}, O103:H25). The *eae* subtype of strain CB10528 (O172:H25) could not be determined. Distribution of the genetic markers *espK*, *espV*, Z2098, and *ureD* in the 55 EHEC strains is shown in Table 1. Overall, the genetic markers investigated were detected in most of the EHEC strains examined. With the exception of the new O26:H11 *stx2*-positive strain (CB14699) and one strain of serotype O157:H7 (CDC-C984), all of the strains were positive for *espK*. The *espV* gene was not detected in two *E. coli* O26:H11 strains (ATCC2196 and CB14699), in O45:H2, O118:H16, O5, or in O123:H11 isolates. The Z2098 gene marker tested negative in only a few strains: the new O26:H11 *stx2* positive clone (strain CB14699) and in the O55:H7 strains. The *ureD* gene was absent in the new O26:H11 *stx2* positive clone (strain CB14699) and in the O55:H7 strains. It was also absent from a few O103:H2 isolates (strains PMK5, CB12062, CB12092). In summary, all of the strains were positive for one or more of the genes *espK*, *espV*, Z2098, and *ureD*, with the exception of the new O26:H11 *stx2* positive clone (CB14699).

Screening Beef Enrichments for *stx1*, *stx2*, *eae*, *espK*, *espV*, Z2098, and *ureD*

A set of 1739 beef samples was screened for the presence of *stx1*, *stx2*, *eae*, *espK*, *espV*, Z2098, and *ureD*. The *stx* genes were detected in 27.95% of the samples (486/1739). The *eae* gene was detected in 21.85% of the samples (380/1739). The two genes were simultaneously present in 10.35% of the samples (180/1739). The genes *espK* and/or Z2098 were detected in 7.42% of the samples (129/1739) (Figure 1A), while *espK* and/or *espV* were found in 130 samples (7.48%) (Figure 1B) and *espK* and/or *ureD* was recorded in 145 samples (8.34%) (Figure 1C).

By using the *stx* and *eae* genes for screening beef samples, following the ISO/CEN TS13136 and MLG5B.05 methods, 180 samples (10.35%) were recorded as *stx/eae* positive and should, therefore, be subjected to a second screening step for EHEC-serogroups. Pre-screening of *stx/eae* positive samples for *espK* with either Z2098 (alternate method A), or *ureD* (alternate method C), provided a 60% reduction of the number of samples that should be submitted to a second screening targeting the O-group gene markers ($n = 71$). Using the alternate method B (*stx/eae/espK/espV*), 80 of the 1739 samples (4.6%) needed to be submitted to a further screening for serogroup determination,



which represents a reduction by 55% of the number of samples subjected to a second screening.

Figure 2 shows the comparison of the alternative methods A–C with the 180 beef samples that tested positive for both the *stx* and *eae* genes. A total of 90 *stx* and *eae* positive beef samples tested negative for *espK*, *espV*, *Z2098*, and *ureD* (sector 8, **Figure 2**). However, the inclusion of the CRISPR_{O26:H11} PCR revealed 12 *espK*, *espV*, *Z2098*, and *ureD* negative samples that were positive for both the new CRISPR_{O26:H11} clone and *eae*-beta, the variant of the intimin gene carried by EHEC O26:H11 (see below).

Screening Beef Samples for *stx1*, *stx2*, *eae*, and CRISPR_{O26:H11}

The 180 beef samples that are positive for *stx* and *eae* were also tested by the CRISPR_{O26:H11} PCR test (SP_O26-E, as described in Delannoy et al., 2015a), that detects the new EHEC O26:H11 French clone (*stx2* and *eae* positive, *espK*, *espV*, *Z2098*, *ureD* negative). Among the 180 samples tested, 20 *stx/eae* positive samples were also found to be positive for SP_O26-E and should therefore be submitted to a further screening for serogroup determination. Interestingly, most of them (16/20) were found positive for *stx2*, 2 samples were positive for *stx1* only and 2 had an unknown *stx* subtype. Twelve of these 20 were also negative for *espK*, *espV*, *Z2098*, and *ureD* (**Figure 2**, sector 8). Finally, when combining the use of CRISPR_{O26:H11} (SP_O26-E PCR test) with alternate method A, 85 samples should be submitted to a second screening targeting the O-group gene markers. When using alternate methods B and C with the CRISPR_{O26:H11} PCR test, 92 and 86 samples should be submitted to the second step, respectively.

Screening Beef Samples for *stx1*, *stx2*, *eae*, and the Top Seven EHEC Serogroups

As recommended in the ISO/TS 13136 (EU) and MLG5B.05 (US) reference methods, the 180 *stx* and *eae* positive samples

were tested for the top seven EHEC serogroups. Among these, 115 samples were positive for at least one of the top seven US regulated EHEC serogroups and 99 were positive for at least one of the top five EHEC serogroups screened by the European ISO/TS 13136 method (data not shown). The most frequently found serogroup was O103 ($n = 71$), followed by O26 ($n = 45$), O121 ($n = 38$), O157 ($n = 18$), O45 ($n = 14$), O145 ($n = 6$), and O111 ($n = 1$). Interestingly, 30 samples were positive for 2 serogroups, 11 were positive for 3 serogroups, and 7 for more than 4 serogroups. In final, 41.74% (48/115) of the beef samples tested positive for more than one O-group marker.

Screening Beef Samples for *stx1*, *stx2*, *eae*, and the *eae* Subtypes

The 180 *stx* and *eae* positive beef samples were further tested for the *eae* subtypes gamma, beta, epsilon and theta, which are associated with one or more of the top seven EHEC serogroups. Among these 180 *stx* and *eae* positive samples, 135 tested positive for at least one of the four *eae*-subtypes: gamma, beta, epsilon and theta (data not shown). The most frequently detected *eae*-subtypes were *eae*-beta ($n = 94$) and *eae*-theta ($n = 65$), followed by *eae*-epsilon ($n = 15$), and *eae*-gamma ($n = 5$). Trying to correlate the *eae* subtype with the serogroup, we identified 51 beef samples for which at least one serogroup was associated with the corresponding related *eae*-subtype. Among these 51 samples, 6 were positive with multiple serogroups.

Comparison of Alternative Methods A–C for Screening Beef Samples

We identified 62 samples that were recorded positive with the three alternative methods A–C (sector 3, **Figure 2**) and therefore must be submitted to a second screening targeting the O-group gene markers. These samples are strongly suspected to contain typical EHEC and five of these are also suspected to contain the new CRISPR_{O26:H11} clone. In addition we found 28 samples (sectors 1, 2, 4, 5, 6, 7 from **Figure 2**) that were positive by one or

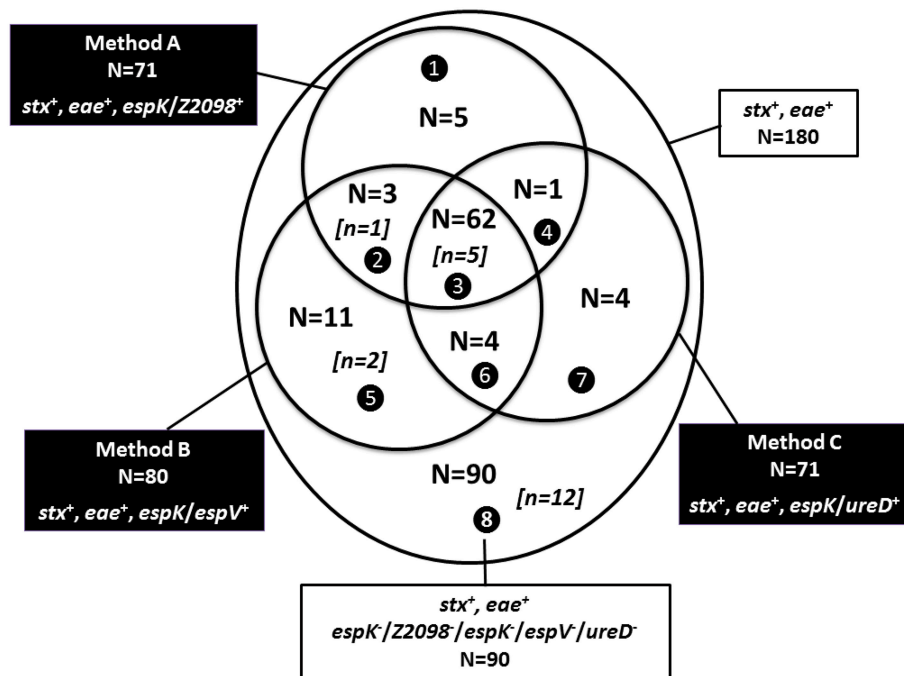


FIGURE 2 | Comparison of Methods A–C on 180 beef samples that tested positive for both *stx* and *eae* genes. *stx*⁺ for samples giving a positive result for *stx1* and/or *stx2*; *eae*⁺ for samples giving a positive result for *eae*; *espK/Z2098*⁺ for samples giving a positive result for *espK* and/or *Z2098*; *espK/espV*⁺ for samples giving a positive result for *espK* and/or *espV*; *espK/ureD*⁺ for samples giving a positive result for *espK* and/or *ureD*. Sector 1 (*stx*⁺, *eae*⁺, *Z2098*⁺), sector 2 (*stx*⁺, *eae*⁺, *Z2098*⁺, *espV*⁺), sector 3 (*stx*⁺, *eae*⁺, *espK/Z2098*⁺, *espK/espV*⁺, *espK/ureD*⁺), sector 4 (*stx*⁺, *eae*⁺, *Z2098*⁺, *ureD*⁺), sector 5 (*stx*⁺, *eae*⁺, *espV*⁺), sector 6 (*stx*⁺, *eae*⁺, *ureD*⁺) and sector 7 (*stx*⁺, *eae*⁺). [n = x] is the number of samples that tested positive by the CRISPR_{O26:H11} PCR assay detecting the new O26 clone, N = x is the total number of samples per sector.

two alternative methods only. PCR results obtained for these 28 samples for *eae* subtypes and top seven O-groups were as follows: only one sample among the 5 samples of sector 1 (Figure 2) was positive for the association of O26 and *eae*-beta, but it tested negative for the different CRISPR_{O26:H11} assays targeting EHEC O26:H11 [SP_O26-C, SP_O26-D, and SP_O26-E, as described in Delannoy et al., 2012a, 2015a (data not shown)]. The only sample from sector 4 (Figure 2) and each of the four samples from sectors 6 and 7 (Figure 2) were found negative by the association of the top seven serogroups and the corresponding *eae*-subtypes. From sector 2 (Figure 2), only one out of three samples was found positive for the new CRISPR_{O26:H11} clone (this sample tested positive by PCR for O26, *eae*-beta, SP_O26-E; and was also *stx2* positive which is consistent with the new clone). The two other samples were not suspected as “presumptive positive” based on the *eae*-subtypes and top seven EHEC serotypes determination. In sector 5 (Figure 2), out of eleven samples three were found “presumptive positive” for EHEC O26 (among them two were suspected to be positive for the new O26:H11 clone). Finally, “presumptive positive” samples were recorded in sectors 2, 3, and 5 which lead to consider the alternate method B as the best one among the other alternate methods for screening EHEC strains. In order to complete the screening of EHEC and not to exclude the new virulent O26 clone, the alternate method B should include the screening of the new CRISPR_{O26:H11} assay.

DISCUSSION

A STEC seropathotype classification has been based upon the serotype association with human epidemics, bloody diarrhea, and HUS, and has been developed as a tool to assess the clinical and public health risks associated with non-O157 EHEC and STEC strains (Karmali et al., 2003). This approach has been of considerable value in defining pathogenic STEC serotypes of importance in cases of human infection (EFSA, 2007; Coombes et al., 2011); however it does not resolve the underlying problem with strains that have not yet been fully serotyped. Furthermore, classification based solely on the presence of seropathotype is inadequate with illnesses linked to STEC serotypes other than O157:H7 that are on the rise worldwide, indicating that some of these organisms may be emerging pathogens. In 2013, the Panel on Biological Hazards (BIOHAZ) of the European Food Safety Authority (EFSA) published a Scientific Opinion on “VTEC-seropathotype and scientific criteria regarding pathogenicity assessment” (EFSA, 2013). This document has focused attention on the applicability of the Karmali seropathotype concept. The document does not provide a scientific definition of a pathogenic STEC but states that the seropathotype classification of Karmali et al. (2003) does not define pathogenic STEC nor does it provide an exhaustive list of pathogenic serotypes. It is not possible to fully define human pathogenic STEC or identify factors for STEC that absolutely predict the potential to cause human disease, but

TABLE 2 | Primer and probe sequences used in this study.

Primer or probe ^a	Sequence 5' → 3' ^b
<i>stx1</i> _F ^c	TTTGTACTGTSACAGCWGAAGCYTTACG
<i>stx1</i> _R ^c	CCCCAGTTCARWGTRAGRTCMACRTC
<i>stx1</i> _P ^c	CTGGATGATCTCAGTGGGCGTTCTTATGTAA
<i>stx2</i> _F ^c	TTTGTACTGTSACAGCWGAAGCYTTACG
<i>stx2</i> _R ^c	CCCCAGTTCARWGTRAGRTCMACRTC
<i>stx2</i> _P ^c	TCGTCAGGCACTGTCTGAAACTGCTCC
<i>eae</i> _F ^d	CATTGATCAGGATTTTTCTGGTGATA
<i>eae</i> _R ^d	CTCATGCGGAAATAGCCGTTA
<i>eae</i> _P ^d	ATAGTCTCGCCAGTATTCGCCACCAATACC
<i>eae</i> -beta_F ^d	GGTGATAATCAGAGTGCACATACA
<i>eae</i> -beta_R ^d	GGCATCAAAATACGTAACCTCGAGTAT
<i>eae</i> -beta_P ^d	CCACAGCAATTACAATACTACCCGGTGCA
<i>eae</i> -gamma_F ^d	GACTGTTAGTGCGACAGTCAGTGA
<i>eae</i> -gamma_R ^d	TTGTTGTCAATTTTCAGTTCATCAAA
<i>eae</i> -gamma_P ^d	TGACCTCAGTCGCTTTAACCTCAGCC
<i>eae</i> -epsilon_F ^d	ATACCCAAATTGTGAAAACGGATA
<i>eae</i> -epsilon_R ^d	CACCTAACACAGCATTACCTGCAA
<i>eae</i> -epsilon_P ^d	CCAGATGTCAGTTTTACCGTAGCCCTACCA
<i>eae</i> -theta_F ^d	TGTTAAAGCACCTGAGGTTACATTTT
<i>eae</i> -theta_R ^d	TCACCAGTAACGTTCTTACCAAGAA
<i>eae</i> -theta_P ^d	TCAACCTTGTTGTCAATTTTCAGTCCATCA
<i>espK</i> _F ^e	GCAGRCATCAAAAGCGAAATCACACC
<i>espK</i> _R ^e	TCGTTTGGTAAGTGTGGCAGATACTC
<i>espK</i> _P ^e	ATTCAGATAGAAGAAGCGCGGGCCAG
<i>espV</i> _F ^e	TCAGGTTCTCGTCTGATGCCGC
<i>espV</i> _R ^e	CTGGTTACAGGCCCTGGAGCAGTCC
<i>espV</i> _P ^e	CTTGCAACACGTTACGCTGCCGAGTATT
Z2098_F ^f	CTGAAAAGAGCCAGAACGTGC
Z2098_R ^f	TGCCCTAAGATCATTACCCGGAC
Z2098_P ^f	TAAGTGTATACCTCCGCGCCG
<i>ureD</i> _F ^g	GCAATAATTGACTCTGATTGCC
<i>ureD</i> _R ^g	GCTGCTGCGGTAAATTTACT
<i>ureD</i> _P ^g	TACGCTGATCACCATGCCCTGGTGC
SP_O26-E_F ^g	AAACCGATCTCCTCATCCTC
SP_O26-E_R ^h	ATCAACATGCAGCGCGAAGC
SP_O26-E_P ^g	CCAGCTACCGACAGTAGTGTGTTCC
<i>rfbE</i> _{O157} -F ^c	TTTCACACTTATTGGATGGTCTCAA
<i>rfbE</i> _{O157} -R ^c	CGATGAGTTTATCTGCAAGGTGAT
<i>rfbE</i> _{O157} -P ^c	AGGACCGCAGAGGAAAGAGAGGAATTAAGG
<i>wzx</i> _{O26} -F ^c	CGCGACGGCAGAGAAAAAT
<i>wzx</i> _{O26} -R ^c	AGCAGGCTTTTATATTCTCAACTTT
<i>wzx</i> _{O26} -P ^c	CCCCGTAAATCAATACTATTTACGAGGTTGA
<i>wzx</i> _{O103} -F ⁱ	CAAGGTGATTACGAAATGCATGT
<i>wzx</i> _{O103} -R ⁱ	GAAAAAAGCACCCCGTACTTAT
<i>wzx</i> _{O103} -P ^j	CATAGCCTGTTGTTTTAT
<i>wbd</i> _{O111} -F ^c	CGAGGCAACACATTATATAGTGCTTT
<i>wbd</i> _{O111} -R ^c	TTTTGAATAGTTATGAACATCTTGTTAGC
<i>wbd</i> _{O111} -P ^c	TTGAATCTCCAGATGATCAACATCGTGAA
<i>wzx</i> _{O121} -F ^j	TGGTCTCTTAGACTTAGGGC
<i>wzx</i> _{O121} -R ^j	TTAGCAATTTCTGTAGTCCAGC
<i>wzx</i> _{O121} -P ^j	TCCAACAATTGGTCGTGAAACAGCTCG

(Continued)

TABLE 2 | Continued

Primer or probe ^a	Sequence 5' → 3' ^b
<i>wzx</i> _{O45} -F ^j	TACGCTGGCTGCAGGG
<i>wzx</i> _{O45} -R ^j	ACTTGCAGCAAAAAATCCCC
<i>wzx</i> _{O45} -P ^j	TTCGTTGCGTTGTGCATGGTGGC
<i>wzy</i> _{O145} -F ^k	ATATTGGGCTGCCACTGATGGGAT
<i>wzy</i> _{O145} -R ^k	TATGGCGTACAATGCACCGCAAAAC
<i>wzy</i> _{O145} -P ^k	AGCAGTGGTTGCGGCACAGCATGGT

^aF, forward primer; R, reverse primer; P, probe.^bAll probes were labeled with 6-HEX or 6-FAM and BHQ1 (Black Hole Quencher).^cOligonucleotide described by Perelle et al. (2004).^dOligonucleotide described by Nielsen and Andersen (2003).^eOligonucleotide described by Delannoy et al. (2013a).^fOligonucleotide described by Delannoy et al. (2013b).^gOligonucleotide described by Delannoy et al. (2012a).^hOligonucleotide described by Delannoy et al. (2015a).ⁱOligonucleotide described by Perelle et al. (2005).^jOligonucleotide described by Bugarel et al. (2010).^kOligonucleotide described by Fratamico et al. (2009).

strains positive for Shiga-toxin (in particular the *stx2* genes) and *eae* (intimin production) genes are associated with a higher risk of more severe illness than other virulence factor combinations (EFSA, 2013). Severe disease, and particularly HUS, is linked to certain serotypes and strains and this link must be the result of particular genetic factors or combinations of factors that have to be determined. A new molecular classification scheme has been proposed by the EFSA Panel on Biological Hazards (BIOHAZ) that relies more on virulence factors than seropathotypes. It is proposed that STEC serogroups O157, O26, O103, O145, O111, and O104 in combination with *stx* and *eae* or *stx* and both *aaiC* (secreted protein of EAEC) and *aggR* (plasmid-encoded regulator) genes should be considered as presenting a potentially higher risk for bloody diarrhea and HUS, such strains are categorized in group I (EFSA, 2013). For any other serogroups in combination with the same genes, the potential risk is regarded as high for diarrhea, but currently unknown for HUS, such strains are categorized in group II. The inclusion of *aaiC* and *aggR* genes in the proposed molecular approach is due to the O104:H4 outbreak, which was caused by a highly virulent strain (Frank et al., 2011). This appears to be an exceptional event (Prager et al., 2014) and future surveillance will provide data that may be used to review the inclusion of these virulence factors but a recent study showed that French cattle are not a reservoir of the highly virulent enteroaggregative Shiga toxin-producing *E. coli* of serotype O104:H4 (Auvray et al., 2012).

Following the EFSA opinion, several laboratories have attempted to develop detection and identification methods for strains of groups I and II, and although substantial progress has been made, a practical method of pathogenic STEC detection has yet to be validated. Molecular methods for screening EHEC O157 and non-O157 in beef products rely currently on the molecular detection of *stx*, *eae*, and the top five or top seven EHEC serogroups in mixed bacterial enrichments as described in the ISO/TS 13136 (EU) and MLG5B.05 (US) reference methods (ISO, 2012; USDA-FSIS, 2014), followed by attempted isolation of the correct strain. These approaches bear the disadvantage that

food samples that carry mixtures of *stx*-negative *E. coli* carrying *eae* genes together with *eae*-negative STEC falsely indicate the presence of EHEC when analyzed for the genes mentioned above (Beutin et al., 2009; Fratamico and Bagi, 2012; Wasilenko et al., 2014). Hence, screening food enrichment broths for *stx* and *eae* genes may cause needless disruption and costs for food producers when the risk of low-virulence STEC is overestimated. The erroneous identification of cucumbers as the source in the outbreak in Germany in 2011 cost European fruit and vegetable producers approximately € 812 million (Commission of the European Communities, 2011). Thus, new DNA targets that unambiguously identify typical EHEC strains (*stx*-positive and *eae*-positive *E. coli* strains) in complex samples are a desirable goal. This new panel of genes might include novel genetic markers recently identified as highly associated to pathogenic STEC strains (Delannoy et al., 2013a,b). Thus, the associations of *espK* with either *espV*, *ureD*, or *Z2098* were found to be the best combinations for more specific and sensitive detection of the top seven EHEC strains, allowing detection of 99.3–100% of these strains. In addition, detection of 93.7% of the typical EHEC strains belonging to other serotypes than the top seven offered the possibility for identifying new emerging typical EHEC strains (Delannoy et al., 2013a,b). Conversely, these different combinations of genetic markers were very rarely associated with STEC (1.6–3.6%) and with non-pathogenic *E. coli* (1.1–3.4%). The objective of the present study was to refine the EHEC screening systems for testing beef samples via the incorporation of these additional gene targets *espK*, *espV*, *ureD*, and *Z2098* in the detection scheme. In addition to these four targets, we included a CRISPR_{O26:H11} target in the detection scheme that has been designed for detecting a new clone of STEC O26:H11 harboring *stx2* only and strongly associated with HUS in France (Delannoy et al., 2015a,b). This new EHEC O26:H11 French clone is positive for *eae* (*eae*-beta subtype) but does not contain any of the above markers *espK*, *espV*, *ureD*, or *Z2098*. The fact that the new French clone is missing *espK*, *espV*, *ureD*, and *Z2098* does not mean that these genes are not required EHEC virulence factors. We have previously demonstrated that they are significantly associated with EHEC strains and not the other pathotypes. It does suggest however that the new French clone may harbor a different set of virulence genes (just like atypical EHEC lack *eae*), which can be investigated through additional genomic studies. In the meantime, although CRISPR sequences are not virulence factors *per se*, we have demonstrated that certain specific spacers are associated with EHEC strains from the top7 serotypes and thus can provide specific and sensitive detection of the top 7 EHEC strains (Delannoy et al., 2012a,b).

To validate the pertinence of this new approach we screened 1739 beef samples and collected 180 samples that tested positive for both *stx* and *eae*, and must be subjected to a second screening step for serogroup determination according to the ISO/TS 13136 (EU) and MLG5B.05 (US) reference methods (ISO, 2012; USDA-FSIS, 2014). Among these 180 samples, 135 tested positive for at least one of the four *eae*-subtypes gamma,

beta, epsilon and theta, which are related to typical EHEC and in particular to those of the top seven serogroups (Bugarel et al., 2010). Introduction of the *eae*-subtypes in the screening step provided a reduction by 25% of the number of *stx/eae* positive samples that should be subjected to a further screening for serogroup determination. A more significant refinement of the first EHEC screening step was achieved by including *espK*, *espV*, *ureD*, *Z2098*, and CRISPR_{O26:H11} target genes in the detection scheme. Thus, a reduction by 52.78–52.22% of the number of samples subjected to a further screening for serogroup determination was obtained by using respectively the alternate method A (*stx/eae/espK/Z2098*) or method C (*stx/eae/espK/ureD*) in combination with the CRISPR_{O26:H11} PCR assay detecting the new O26 clone. A reduction by 48.88% of the number of “presumptive positive” samples was obtained using the alternate method B (*stx/eae/espK/espV*) in association with the CRISPR_{O26:H11} PCR assay. Given the additional information on the association of the top seven serogroups and the *eae*-subtypes, we determined the last approach, i.e., method B (*stx/eae/espK/espV*) in association with the CRISPR_{O26:H11} PCR assay, as the best approach to narrow down the EHEC screening step in beef samples. Using such an approach, 92 samples must be subjected to a further screening for serogroup determination vs. 180 with the conventional *stx/eae* approach used in the ISO/TS 13136 (EU) and MLG5B.05 (US) reference methods (ISO, 2012; USDA-FSIS, 2014). This constitutes a significant reduction (almost 50%) of the number of samples subjected to a second screening targeting the O-group gene markers. Moreover, this approach is in line with the EFSA opinion that has identified STEC strains of groups I and II as presenting the potential higher risk for diarrhea and HUS (EFSA, 2013). Identification of additional gene markers, i.e., *espK*, *espV*, and CRISPR_{O26:H11} to better distinguish typical EHEC from other *E. coli* pathogroups would substantially enhance the power of EHEC test systems providing a significant reduction of “presumptive positive” in beef samples. Such a new approach would provide to the agroindustry a novel method for tracking EHEC in food samples. This work should be considered with interest to draw up the outline of a future standard that will follow the recommendations of EFSA.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: SD, PF. Performed the experiments: SD, BC, SI, HW, JD, IB. Analyzed the data: SD, BC, PF. Contributed reagents/materials/analysis tools: SD, LB, JD, IB, PF. Wrote the paper: SD, PF. Critical revision of the paper for important intellectual content: SD, BC, SI, HW, LB, JD, IB, PF.

FUNDING

The project was partially financed by the French “joint ministerial program of R&D against CBRNE risks” (Grant number C17609-2).

REFERENCES

- Auvray, F., Dilasser, F., Bibbal, D., Kérourédan, M., Oswald, E., and Brugère, H. (2012). French cattle is not a reservoir of the highly virulent enteroaggregative Shiga toxin-producing *Escherichia coli* of serotype O104:H4. *Vet Microbiol.* 158, 443–445. doi: 10.1016/j.vetmic.2012.02.029
- Beutin, L., and Fach, P. (2014). Detection of shiga toxin-producing *Escherichia coli* from nonhuman sources and strain typing. *Microbiol. Spectr.* 2:EHEC-0001-2013. doi: 10.1128/microbiolspec.EHEC-0001-2013
- Beutin, L., Jahn, S., and Fach, P. (2009). Evaluation of the 'GeneDisc' real-time PCR system for detection of enterohaemorrhagic *Escherichia coli* (EHEC) O26, O103, O111, O145 and O157 strains according to their virulence markers and their O- and H-antigen-associated genes. *J. Appl. Microbiol.* 106, 1122–1132. doi: 10.1111/j.1365-2672.2008.04076.x
- Bibbal, D., Loukiadis, E., Kérourédan, M., Peytavin de Garam, C., Ferré, F., Cartier, P., et al. (2014). Intimin gene (*eae*) subtype-based real-time PCR strategy for specific detection of Shiga toxin-producing *Escherichia coli* serotypes O157:H7, O26:H11, O103:H2, O111:H8, and O145:H28 in cattle feces. *Appl. Environ. Microbiol.* 80, 1177–1184. doi: 10.1128/AEM.03161-13
- Bielaszewska, M., Mellmann, A., Bletz, S., Zhang, W., Köck, R., Kossow, A., et al. (2013). Enterohemorrhagic *Escherichia coli* O26:H11/H-: a new virulent clone emerges in Europe. *Clin. Infect. Dis.* 56, 1373–1381. doi: 10.1093/cid/cit055
- Brooks, J. T., Sowers, E. G., Wells, J. G., Greene, K. D., Griffin, P. M., Hoekstra, R. M., et al. (2005). Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. *J. Infect. Dis.* 192, 1422–1429. doi: 10.1086/466536
- Bugarel, M., Beutin, L., and Fach, P. (2010). Low-density microarray targeting non-locus of enterocyte effacement effectors (*nle* genes) and major virulence factors of Shiga toxin-producing *Escherichia coli* (STEC): a new approach for molecular risk assessment of STEC isolates. *Appl. Environ. Microbiol.* 76, 203–211. doi: 10.1128/AEM.01921-09
- Commission of the European Communities (2011). *Commission Staff Working Document on Lessons Learned from the 2011 Outbreak of Shiga Toxin-Producing Escherichia coli (STEC) O104:H4 in Sprouted Seeds*. SANCO/13004/2011, 1–23. Available online at http://ec.europa.eu/food/food/biosafety/salmonella/docs/cswd_lessons_learned_en.pdf (Accessed on August 8, 2015).
- Coombes, B. K., Gilmour, M. W., and Goodman, C. D. (2011). The evolution of virulence in non-O157 shiga toxin-producing *Escherichia coli*. *Front. Microbiol.* 2:90. doi: 10.3389/fmicb.2011.00090
- Delannoy, S., Beutin, L., Burgos, Y., and Fach, P. (2012b). Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the CRISPR locus as a target for a diagnostic real-time PCR. *J. Clin. Microbiol.* 50, 3485–3492. doi: 10.1128/JCM.01656-12
- Delannoy, S., Beutin, L., and Fach, P. (2012a). Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. *J. Clin. Microbiol.* 50, 4035–4040. doi: 10.1128/JCM.02097-12
- Delannoy, S., Beutin, L., and Fach, P. (2013a). Discrimination of enterohemorrhagic *Escherichia coli* (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. *J. Clin. Microbiol.* 51, 3257–3262. doi: 10.1128/JCM.01471-13
- Delannoy, S., Beutin, L., and Fach, P. (2013b). Towards a molecular definition of enterohemorrhagic *Escherichia coli* (EHEC): detection of genes located on O island 57 as markers to distinguish EHEC from closely related enteropathogenic *E. coli* strains. *J. Clin. Microbiol.* 51, 1083–1088. doi: 10.1128/JCM.02864-12
- Delannoy, S., Mariani-Kurkdjian, P., Bonacorsi, S., Liguori, S., and Fach, P. (2015a). Characteristics of emerging human-pathogenic *Escherichia coli* O26:H11 strains isolated in France between 2010 and 2013 and carrying the *stx2d* gene only. *J. Clin. Microbiol.* 53, 486–492. doi: 10.1128/JCM.02290-14
- Delannoy, S., Mariani-Kurkdjian, P., Bonacorsi, S., Liguori, S., Ison, S. A., and Fach, P. (2015b). Draft genome sequences of human-pathogenic *Escherichia coli* O26:H11 strains carrying the *stx2* gene only and circulating in France. *Genome Announc.* 3:e00852-15. doi: 10.1128/genomeA.00852-15
- EFSA (European Food Safety Authority) (2007). Scientific Opinion of the Panel on Biological Hazards (BIOHAZ) on monitoring of verotoxigenic *Escherichia coli* (VTEC) and identification of human pathogenic VTEC types. *EFSA J.* 5, 79, 1–61. doi: 10.2903/j.efsa.2007.579
- EFSA (European Food Safety Authority) (2013). Scientific Opinion of the Panel on Biological Hazards (BIOHAZ) on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment. *EFSA J.* 11, 3138, 1–106. doi: 10.2903/j.efsa.2013.3138
- EFSA (European Food Safety Authority, European Centre for Disease Prevention and Control) (2014). The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012. *EFSA J.* 12, 3547, 312. doi: 10.2903/j.efsa.2014.3547
- Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., an der Heiden, M., et al. (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* 365, 1771–1780. doi: 10.1056/NEJMoa1106483
- Fratamico, P. M., and Bagi, L. K. (2012). Detection of Shiga toxin-producing *Escherichia coli* in ground beef using the GeneDisc real-time PCR system. *Front. Cell Infect. Microbiol.* 2:152. doi: 10.3389/fcimb.2012.00152
- Fratamico, P. M., DebRoy, C., Miyamoto, T., and Liu, Y. (2009). PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E. coli* O145 O-antigen gene cluster and the shiga toxin 1 and shiga toxin 2 genes. *Foodborne Pathog. Dis.* 6, 605–611. doi: 10.1089/fpd.2008.0254
- Gould, L. H., Mody, R. K., Ong, K. L., Clogher, P., Cronquist, A. B., Garman, K. N., et al. (2013). Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000–2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Pathog. Dis.* 10, 453–460. doi: 10.1089/fpd.2012.1401
- ISO (International Organization for Standardization) (2012). *ISO/TS 13136:2012, Microbiology of Food and Animal Feed—Real-Time Polymerase Chain Reaction (PCR) –Based Method for the Detection of Food-Borne Pathogens—Horizontal Method for the Detection of Shiga Toxin-Producing Escherichia coli (STEC) and the Determination of O157, O111, O26, O103, and O145 Serogroups*. Available online at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53328. (Accessed on August 8, 2015).
- Johnson, K. E., Thorpe, C. M., and Sears, C. L. (2006). The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. *Clin. Infect. Dis.* 43, 1587–1595. doi: 10.1086/509573
- Karmali, M. A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., et al. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *J. Clin. Microbiol.* 41, 4930–4940. doi: 10.1128/JCM.41.11.4930-4940.2003
- Livezey, K. W., Groschel, B., and Becker, M. M. (2015). Use of the *ecf1* gene to detect Shiga toxin-producing *Escherichia coli* in beef samples. *J. Food Prot.* 78, 675–684. doi: 10.4315/0362-028X.JFP-14-417
- Madic, J., Peytavin de Garam, C., Vingadassalon, N., Oswald, E., Fach, P., Jamet, E., et al. (2010). Simplex and multiplex real-time PCR assays for the detection of flagellar (H-antigen) *fliC* alleles and intimin (*eae*) variants associated with enterohaemorrhagic *Escherichia coli* (EHEC) serotypes O26:H11, O103:H2, O111:H8, O145:H28 and O157:H7. *J. Appl. Microbiol.* 109, 1696–1705. doi: 10.1111/j.1365-2672.2010.04798.x
- Madic, J., Vingadassalon, N., Peytavin de Garam, C. P., Marault, M., Scheutz, F., Brugère, H., et al. (2011). Detection of Shiga toxin-producing *Escherichia coli* serotypes O26:H11, O103:H2, O111:H8, O145:H28, and O157:H7 in raw-milk cheeses by using multiplex real-time PCR. *Appl. Environ. Microbiol.* 77, 2035–2041. doi: 10.1128/AEM.02089-10
- Michelet, L., Delannoy, S., Devillers, E., Umhang, G., Aspan, A., Juremalm, M., et al. (2014). High-throughput screening of tick-borne pathogens in Europe. *Front. Cell Infect. Microbiol.* 4:103. doi: 10.3389/fcimb.2014.00103
- Miko, A., Delannoy, S., Fach, P., Strockbine, N. A., Lindstedt, B. A., Mariani-Kurkdjian, P., et al. (2013). Genotypes and virulence characteristics of Shiga toxin-producing *Escherichia coli* O104 strains from different origins and sources. *Int. J. Med. Microbiol.* 303, 410–421. doi: 10.1016/j.ijmm.2013.05.006
- Nielsen, E. M., and Andersen, M. T. (2003). Detection and characterization of verocytotoxin-producing *Escherichia coli* by automated 5' nuclease PCR assay. *J. Clin. Microbiol.* 41, 2884–2893. doi: 10.1128/JCM.41.7.2884-2893.2003
- Oswald, E., Schmidt, H., Morabito, S., Karch, H., Marchès, O., and Caprioli, A. (2000). Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. *Infect. Immun.* 68, 64–71. doi: 10.1128/IAI.68.1.64-71.2000

- Perelle, S., Dilasser, F., Grout, J., and Fach, P. (2004). Detection by 5'-nuclease PCR of Shiga-toxin producing *Escherichia coli* O26, O55, O91, O103, O111, O113, O145 and O157:H7, associated with the world's most frequent clinical cases. *Mol. Cell Probes*. 18, 185–192. doi: 10.1016/j.mcp.2003.12.004
- Perelle, S., Dilasser, F., Grout, J., and Fach, P. (2005). Detection of *Escherichia coli* serogroup O103 by real-time polymerase chain reaction. *J. Appl. Microbiol.* 98, 1162–1168. doi: 10.1111/j.1365-2672.2005.02545.x
- Prager, R., Lang, C., Aurass, P., Fruth, A., Tietze, E., and Flieger, A. (2014). Two novel EHEC/EAEC hybrid strains isolated from human infections. *PLoS ONE* 9:e95379. doi: 10.1371/journal.pone.0095379
- USDA-FSIS (2014). *Detection and Isolation of non-O157 Shiga Toxin-Producing Escherichia coli (STEC) from Meat Products and Carcass and Environmental Sponges*. In Microbiological Laboratory Guidelines chapter 5B.05. Available online at <http://www.fsis.usda.gov/wps/wcm/connect/7ffc02b5-3d33-4a79-b50c-81f208893204/MLG-5B.pdf?MOD=AJPERES> (Accessed on December 10, 15).
- Wang, F., Yang, Q., Kase, J. A., Meng, J., Clotilde, L. M., Lin, A., et al. (2013). Current trends in detecting non-O157 Shiga toxin-producing *Escherichia coli* in food. *Foodborne Pathog. Dis.* 10, 665–677. doi: 10.1089/fpd.2012.1448
- Wasilenko, J. L., Fratamico, P. M., Sommers, C., DeMarco, D. R., Varkey, S., Rhoden, K., et al. (2014). Detection of Shiga toxin-producing *Escherichia coli* (STEC) O157:H7, O26, O45, O103, O111, O121, and O145, and Salmonella in retail raw ground beef using the DuPont™ BAX® system. *Front. Cell Infect. Microbiol.* 4:81. doi: 10.3389/fcimb.2014.00081
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Delannoy, Chaves, Ison, Webb, Beutin, Delaval, Billet and Fach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Implementation of Whole Genome Sequencing (WGS) for Identification and Characterization of Shiga Toxin-Producing *Escherichia coli* (STEC) in the United States

Rebecca L. Lindsey¹, Hannes Pouseele², Jessica C. Chen³, Nancy A. Strockbine¹ and Heather A. Carleton^{1*}

¹ Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA, ² Applied Maths NV, Sint-Martens-Latem, Belgium, ³ IHRC Inc., Atlanta, GA, USA

OPEN ACCESS

Edited by:

Pina Fratamico,
United States Department of
Agriculture, Agricultural Research
Service, USA

Reviewed by:

Michel Drancourt,
Aix Marseille Université, France
Alan Leonard,
Florida Institute of Technology, USA

*Correspondence:

Heather A. Carleton
hcarleton@cdc.gov

Specialty section:

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

Received: 02 March 2016

Accepted: 06 May 2016

Published: 23 May 2016

Citation:

Lindsey RL, Pouseele H, Chen JC,
Strockbine NA and Carleton HA
(2016) Implementation of Whole
Genome Sequencing (WGS) for
Identification and Characterization of
Shiga Toxin-Producing *Escherichia coli*
(STEC) in the United States.
Front. Microbiol. 7:766.
doi: 10.3389/fmicb.2016.00766

Shiga toxin-producing *Escherichia coli* (STEC) is an important foodborne pathogen capable of causing severe disease in humans. Rapid and accurate identification and characterization techniques are essential during outbreak investigations. Current methods for characterization of STEC are expensive and time-consuming. With the advent of rapid and cheap whole genome sequencing (WGS) benchtop sequencers, the potential exists to replace traditional workflows with WGS. The aim of this study was to validate tools to do reference identification and characterization from WGS for STEC in a single workflow within an easy to use commercially available software platform. Publicly available serotype, virulence, and antimicrobial resistance databases were downloaded from the Center for Genomic Epidemiology (CGE) (www.genomicepidemiology.org) and integrated into a genotyping plug-in with *in silico* PCR tools to confirm some of the virulence genes detected from WGS data. Additionally, down sampling experiments on the WGS sequence data were performed to determine a threshold for sequence coverage needed to accurately predict serotype and virulence genes using the established workflow. The serotype database was tested on a total of 228 genomes and correctly predicted from WGS for 96.1% of O serogroups and 96.5% of H serogroups identified by conventional testing techniques. A total of 59 genomes were evaluated to determine the threshold of coverage to detect the different WGS targets, 40 were evaluated for serotype and virulence gene detection and 19 for the *stx* gene subtypes. For serotype, 95% of the O and 100% of the H serogroups were detected at > 40x and ≥ 30x coverage, respectively. For virulence targets and *stx* gene subtypes, nearly all genes were detected at > 40x, though some targets were 100% detectable from genomes with coverage ≥ 20x. The resistance detection tool was 97% concordant with phenotypic testing results. With isolates sequenced to > 40x coverage, the different databases accurately predicted serotype, virulence, and resistance from WGS data, providing a fast and cheaper alternative to conventional typing techniques.

Keywords: *Escherichia coli*, whole genome sequence, STEC, next generation sequencing, *stx* subtyping, *Escherichia coli* serotypes

INTRODUCTION

Foodborne bacteria pose a major threat to public health. To prevent widespread infections due to these bacteria as well as detect outbreaks, rapid and accurate identification and subtyping of these bacteria is key. Shiga toxin-producing *Escherichia coli* (STEC) is an important foodborne pathogen estimated to cause over 265,100 illnesses each year in the United States (Scallan et al., 2011). STEC may present as a mild gastroenteritis, diarrhea, grossly bloody diarrhea and hemolytic uremic syndrome (HUS), and infection may be fatal. In the United States an estimated 96,500 O157 STEC and 168,690 non-O157 STEC infections occur each year and result in over 3600 hospitalizations and 30 deaths annually (Scallan et al., 2011).

STEC is a nationally reportable disease in the U.S. and clinical laboratory requirements for forwarding the STEC positive isolate or specimen to the public health laboratory vary by state. Once a STEC positive isolate or specimen arrives at the local or state public health laboratories it undergoes further characterization. These isolates are routinely subtyped using pulsed-field gel electrophoresis (PFGE) and submitted to the national surveillance network for foodborne bacteria, PulseNet, as well as characterized using conventional techniques for phenotype, serotype, and virulence. Workflows at public health laboratories for characterization for STEC can vary, current methods for characterization of STEC in the Enteric Diseases Laboratory Branch at the Centers for Disease Control and Prevention include panels of 22–49 phenotypic tests for identification, agglutination assays with 270 pooled and individual O- and H-specific antisera for serotyping (determination of 188 O and 53 H antigens), panels of five to 10 PCR assays for virulence profiling and broth microdilution assays for antimicrobial susceptibility testing. These methods require complex workflows, expensive reagents, labor-intensive quality control procedures, specialized training, and typically take 1–3 weeks to complete. Therefore a need exists to simplify workflows and reduce costs and time associated with subtyping and characterization of STEC, possibly through whole genome sequencing.

Whole genome sequencing (WGS) using benchtop instruments makes WGS possible in a public health lab setting. These machines are relatively easy to operate; the cost per isolate is low; and turnaround time for generating WGS data is within days rather than the 1–3 weeks required for current, conventional methods. Since the serotype, virulence and antimicrobial resistance profile may be predicted from the genome sequence, WGS may replace almost all reference characterization of STEC in the public health laboratory. Additionally, the sequence data also provide a level of strain discrimination and precision that is better than any subtyping method hitherto used for outbreak detection and investigation. Thus, almost all characterization of STEC in the public health laboratory can be replaced by WGS using one single efficient workflow. However, converting the WGS data into interpreted output that is useful for public health professionals is a real challenge.

To address this challenge, The Center for Genomic Epidemiology (CGE) (www.genomicepidemiology.org) has developed a suite of web-based tools for *in silico* analysis of bacterial whole genome sequence (Cosentino et al., 2013; Joensen et al., 2014). These tools include a serotype detection procedure (SerotypeFinder) and resistance and virulence prediction tool (ResFinder and VirulenceFinder) for analysis of *E. coli* and other bacterial WGS data (Zankari et al., 2013; Kleinheinz et al., 2014; Joensen et al., 2015). To characterize an isolate, WGS is uploaded to the website, and depending on the analysis requested, a report of the isolate's serotype, virulence, and resistance gene content is returned within several minutes to hours. Since many WGS analysis tools will accept data of any quality, it is important to understand the data quality requirements for the information being sought to interpret negative results correctly.

Although the CGE tools are useful for analysis in a setting where the isolate throughput is low and data analysis is centralized, for WGS analysis tools to be effective in a public health laboratory setting that processes tens to hundreds of WGS isolate sets per week, all the tools need to be merged into a single platform that performs WGS quality assessment and can also be used with a database that includes sufficient patient and sample information about the isolates to be able to interpret them in the proper epidemiological context, e.g., the outbreak setting. Moreover, the platform needs to be simple and user-friendly so that it may be used by public health professionals with limited bioinformatics skills. While numerous commercial and public domain software are available to analyze WGS data, very few combine databasing, WGS and other analytical capabilities, which are highly desirable in national and international laboratory surveillance networks. One exception is BioNumerics v7.5[®] (Applied Maths, Austin, TX), which is a commercial, customizable WGS quality assessment, analysis, and database software package that may be used for all these purposes. Thus the serotype, virulence, and resistance gene detection tools from CGE and *in silico* PCR tools to confirm results from virulence prediction tools can be integrated into a single push button tool in BioNumerics v7.5[®], the genotyping plug-in. This plug-in can be used in the same database that contains the sequence data for subtyping for surveillance purposes and only requires a *de novo* assembled genome. Therefore, within the one software platform, reference characterization and WGS analysis for outbreak detection can be performed rapidly and requires little bioinformatics training for the user.

In this study, we performed validation of a genotyping plug-in within BioNumerics for identification of O and H genes using a diversity set of nearly 200 genomes. We further demonstrated the utility of the detection tools within the genotyping plug-in on isolates that were sequenced in-house and tested by traditional methods for serotype, virulence gene content, and antimicrobial susceptibility. In addition, we down-sampled the sequence reads of the later set to determine the minimum genome coverage needed for the program to detect the intended target genes and present *in silico* PCR tools to confirm selected results of the virulence detection tool.

MATERIALS AND METHODS

Whole Genome Sequencing and Analysis

Validation of the genotyping plug-in within BioNumerics version 7.5 was performed using whole genome sequences of nearly 300 in-house sequenced and publically available STEC genomes for which conventional serotyping results were also available. The sequence data were generated on PacBio, Roche 454, and Illumina sequencing platforms (Table S1). An additional 17 genomes of *Shigella* and corresponding antimicrobial susceptibility testing data were included to evaluate the resistance finder tool. Moreover, a set of 106 isolates with traditional serotype and virulence PCR results performed in house were sequenced on the Illumina MiSeq or HiSeq sequencer platforms and selected for the down sampling experiment (Table S1). For in-house sequenced isolates, DNA was extracted using the Qiagen Blood and Tissue kit, libraries were prepared using NexteraXT (MiSeq) or NEB Next (HiSeq), and sequenced on the MiSeq or HiSeq using 2×250 bp chemistry. Sequence quality was evaluated on a per genome basis using BioNumerics version 7.5[®]. All genomes passed the basic quality metrics for raw sequence data from Illumina sequencers of average Q-score >30 in both reads and at least 40X average coverage with expected genome size for *E. coli* of 5 Mb. Read files of in house generated sequence data were uploaded to NCBI SRA using BioNumerics v.7.5 NCBI uploader (see Table S1). Genomes were processed through the BioNumerics Calculation Engine for *de novo* assemblies using the wgMLST client plug-in. The assembly was done using SPAdes version 3.5.0 integrated into the wgMLST plug-in and basic assembly metrics were calculated and used for quality assessment.

Analysis Using Genotyping Plug-in

Assembled sequence data was analyzed using the genotyping plug-in. The genotyping plug-in contains databases for serotype, virulence and resistance prediction (consisting of annotated allelic variants for genes encoding serotype, virulence factors and antimicrobial resistance), and for plasmid and prophage detection obtained from the Center for Genomic Epidemiology (DTU, Lyngby, Denmark) (<https://cge.cbs.dtu.dk/services/data.php>). The genotyping plug-in also contains an *in silico* PCR tool for the detection of Shiga toxin gene subtypes and virulence genes using previously published primers (Paton and Paton, 1998; Scheut et al., 2012). The various “finder” tools use a blast-based approach to detect the genes of interest in the *de novo* assembled genome, and subsequently identifies them against the appropriate reference database. Detection parameters were set to 90% sequence identity and 60% sequence coverage. As a quality metric and a guard against blindly extrapolating the serotype, virulence or resistance prediction, for each similarity-based association, a discrimination score is calculated, indicating how good the closest known allele in the respective database fits the sample data with respect to the runner-up allele. The *in silico* PCR tools, mimicking the wet lab PCR process, excise a particular part of the genome, defined by forward and reverse primer pairs. In detecting a primer, at most 1 mismatch was allowed.

Downsampling Analysis

A set of 59 genomes were downsampled. A set of 40 genomes were used to validate the serotype and virulence gene finder, and the remaining set of 19 were used to validate the *stx* subtyper. The genomes were downsampled to 40x, 30x, 20x, and 10x coverage using the Computational Genomics Pipeline (CG-Pipeline; <https://github.com/liskatz/CG-Pipeline>; Katz et al., 2011). Downsampled genomes were assembled and analyzed using the genotyping plug-in BioNumerics v.7.5 with the settings outlined above to determine limit of detection for WGS-based identification tools.

Conventional Testing Procedures

Conventional testing of isolates was completed in the *Escherichia Shigella* reference laboratory at the Centers for Disease Control and Prevention, USA. Serotyping was performed with O- and H-specific antisera from the Statens Serum Institut (Copenhagen, Denmark) by standard methods in a microtiter format (Ewing, 1986). For the virulence genes real-time or conventional PCR for the presence of Shiga toxin 1 and 2 (*stx*₁, *stx*₂), *stx* subtyping (*stx*_{1a}, *stx*_{1c}, *stx*_{1d}, *stx*_{2a}, *stx*_{2c}, *stx*_{2d}, *stx*_{2e}, *stx*_{2f}, and *stx*_{2g}), intimin (*eae*) and hemolysin (*ehxA*) genes was performed (Paton and Paton, 1998; Scheut et al., 2012). Broth microdilution assays to determine antimicrobial susceptibility was done by the National Antimicrobial Resistance Monitoring (NARMS) laboratory using previously published techniques (CDC, 2013). Resistance data were interpreted using Clinical Laboratory Standards Institute criteria (Clinical Laboratory Standards Institute, 2012).

RESULTS

Validation of the Serotype Detection Tool in BioNumerics

The serotype detection tool within BioNumerics was validated on a total of 188 isolates for which WGS data and conventional serotype information was available. These publically available genomes were sequenced by either PacBio, Illumina, or 454 technology (see Table S1). The genomes represent 30 O serogroups and 26 H serogroups for a total of 76 serotypes (Table 1). Several representatives of the top 20 serotypes as well as a representation of a diverse collection of serotypes were selected for this set of genomes.

A total of 29 O and 25 H serogroups were identified from the WGS data of the 30 O and 26 H serogroups detected by conventional methods, one O118 serogroup isolate was not detected by WGS data, and an H47 isolate was typed as an H7 by WGS. Comparisons to the traditional O serogroup results with the predictions from the WGS data showed that 96.3% (181/188) of the O serogroups were accurately predicted from the WGS data. For the H serogroup, 95.9% (164/171) of the H antigens were accurately predicted from the WGS data, H antigen detection for non-motile isolates was not counted since such isolates are non-typable by the phenotypic methods (Table 1). There were only 4 isolates that had a different O serogroup predicted and 6 isolates that had a different H serogroup predicted compared to traditional typing results (see Table S1).

TABLE 1 | Validation of serotype detection tool within the genotyping plug-in on a set of 188 isolates.

Antigen	Number of strains phenotypically determined or reported in the literature with antigen	Number of strains with antigen determined by WGS in agreement with phenotypically determined antigen
O1	3	2 ^b
O2	2	2
O6	10	10
O7	4	4
O8	3	3
O9	3	3
O15	2	2
O16	7	7
O18	3	3
O25	3	3
O26	12	11 ^a
O45	3	3
O55	20	20
O78	2	2
O83	2	2
O91	6	5 ^a
O103	4	4
O104	4	4
O111	23	23
O118	1	0 ^b
O119	1	1
O121	2	2
O127	1	1
O128	16	15 ^a
O145	8	7 ^b
O146	2	2
O149	2	1 ^a
O157	36	35
O165	1	1
O174	3	3
H1	8	8
H2	18	18
H4	8	8
H6	13	13
H7	40	40
H8	5	4 ^b
H9	2	2
H11	13	11 ^a
H12	4	3 ^a
H14	1	1
H16	2	2
H17	1	1
H19	6	6
H20	2	2
H21	16	14 ^a
H25	4	4

(Continued)

TABLE 1 | Continued

Antigen	Number of strains phenotypically determined or reported in the literature with antigen	Number of strains with antigen determined by WGS in agreement with phenotypically determined antigen
H28	4	4
H31	1	1
H34	1	1
H37	1	1
H39	1	1
H43	4	4
H45	5	5
H47	1	0 ^a
H48	7	7
H49	3	3

Number of isolates given for positive by conventional and WGS serotype tests (isolate details listed in Table S1). 17 isolates were non-motile by traditional testing and not counted in the H WGS test results.

^aDiscrepant conventional and WGS serotyping results are noted by isolate in Table S1.

^bAntigen not predicted from WGS data.

These were not sequenced by us but were downloaded from NCBI and it is possible that the data on NCBI may contain errors. However, since we do not have access to these isolates we cannot confirm the WGS results and phenotype. Overall, the serotyper tool predicted the serotype correctly in 94.2% (161/171) of the tested genomes.

Robustness of Serotype and Virulence Gene Prediction in WGS Datasets

To determine the sensitivity of the serotype and virulence gene predictions by WGS, a set of 40 isolates was selected that had all been sequenced by Illumina MiSeq or HiSeq and serotyped and virulence gene characterized using PCR methods. These genomes ranged in coverage from 40x to 267x coverage. Using the serotype detection tool in the genotyping plug-in, all but one of the O serogroups were predicted (95%) (38/40 isolates), both isolates belonging to the O153 serogroup were not predicted (Table 2). For those isolates where no O serogroup was predicted, the genomes ranged from 119x to 153x coverage, suggesting that sequence coverage was not a factor in being able to predict this particular serogroup. All of the H serotypes were predicted correctly when considering motile isolates, i.e. isolates that could be phenotypically verified by agglutination. The sequencing reads per isolate were then randomly down sampled to 40, 30, 20, and 10 times coverage and then analyzed in BioNumerics. These genomes were assembled *de novo* and the serotype and virulence genes predicted. In the down-sampled datasets, at 40x coverage 77.5% of O and 100% of H serogroups were correctly identified. For the remaining 30x, 20x, and 10x coverage levels, O serogroups were predicted correctly in 77.5, 52.5, and 17.5% of isolates and H serogroups were predicted in 100, 95, and 70% of isolates, respectively. The best prediction of O and H serogroup was from genomes at greater than 40x coverage.

For the original sequence and 40x, 30x, and 20x down sampled genomes, there was 100% concordance between the virulence

TABLE 2 | Limit of detection for O and H antigens in a downsampled WGS data set from 40 strains.

Antigen ^b	Number of strains with phenotypically determined antigen	Number of strains by WGS having the phenotypically determined antigen at different sequence coverage levels				
		Original sequence coverage ^a	40x coverage	30x coverage	20x coverage	10x coverage
O5	2	2	2	2	2	1
O26	2	2	2	2	1	0
O45	2	2	2	2	1	1
O69	2	2	2	2	1	0
O71	2	2	2	2	2	2
O76	2	2	2	2	2	1
O80	2	2	2	2	1	0
O91	2	2	2	2	0	0
O103	2	2	2	2	2	0
O104	2	2	2	2	0	0
O111	2	2	2	2	2	0
O113	2	2	2	2	1	0
O118	2	2	1	2	1	0
O121	2	2	2	2	2	2
O145	2	2	0	0	0	0
O146	2	2	1	1	1	0
O153	2	0	0	0	0	0
O157	2	2	2	2	2	0
O165	2	2	0	0	0	0
O174	2	2	2	0	0	0
H2	7	7	7	7	7	6
H4	3	3	3	3	3	2
H7	3	3	3	3	3	2
H8	2	2	2	2	2	2
H11	5	5	5	5	5	4
H14	1	1	1	1	1	1
H16	2	2	2	2	2	2
H19	4	4	4	4	4	2
H21	3	3	3	3	3	2
H25	1	2	2	2	2	1
H28	3	3	3	3	2	2

Strain identifiers listed in Table S1.
^aOriginal coverage ranged from 40 to 267x.
^bSix H serogroups were called from the WGS data that were typed as non-motile by conventional methods and not included here.

detection tool and *in silico* PCR and conventional real-time PCR assay for Shiga toxin 1 and 2 (*stx*₁, *stx*₂), intimin (*eae*), and hemolysin (*ehxA*) genes when a call was made (see **Table 3**). At 10x coverage, few virulence genes were identified from the WGS. The virulence detection databases did not identify *stx*₂ in a STEC O76:H19 isolate which was detected by *in silico* PCR. Additionally, the *in silico* PCR did not identify *stx*₂ in two isolates that were identified as *stx*₂ positive by the virulence detection tool databases. By using both the virulence detection databases and *in silico* PCR, all *stx*₂ positive isolates identified by conventional typing methods were also identified from the WGS at 20x coverage or higher. For the other virulence gene targets, *stx*₁ was detected in all 21 of the isolates positive by conventional testing at ≥ 20x coverage, *eae* in all 26 at ≥ 30 coverage, and

ehxA gene detection from WGS data was 100% concordant in the assemblies from the original sequence read set.

Prediction of *stx* Gene Subtype Using Virulence Gene Database and *In silico* PCR

A total of 19 isolates were examined that had complete conventional *stx* gene subtype results and WGS results. These isolates represented the a, c, and d subtypes of *stx*₁ and the a, b, c, d, e, f, and g subtypes of *stx*₂. All the *stx* subtypes except *stx*_{2c} were detected using the *in silico* PCR tool or the virulence detection database at original coverage levels and down to 10x coverage (see **Table 4**). For *stx*_{2c}, only one of the two isolates identified as positive by conventional laboratory testing was detected as

TABLE 3 | Limit of detection of virulence genes in a down sampled WGS data set from 40 STEC and one hybrid STEC/EAEC O104:H4 by both a blast and *in silico* PCR approach.

Trait	Number of strains with trait by real-time PCR	Number strains determined from WGS data with indicated trait at different sequence coverage levels									
		Original sequence coverage ^a		40x coverage		30x coverage		20x coverage		10x coverage	
		Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR
<i>stx1</i>	27	27	27	27	27	27	27	27	27	25	26
<i>stx2</i>	21	20 ^b	19 ^c	20	19	20	20	20	20	20	18
<i>eae</i>	26	26	26	26	24	26	26	24	20	8	9
<i>ehxA</i>	29	29	29	26	26	25	25	17	20	8	11

Strain identifiers listed in Table S1.

^aOriginal coverage was 40x to 267x.

^bFor the original sequence files, *stx2* was missed in an *E. coli* O76:H19 using the genotyping plug-in that was detected by *in silico* PCR.

^cThe *in silico* PCR did not detect *stx2* in 2 isolates though it was detected by the genotyping plug-in.

TABLE 4 | Limit of detection of *stx* gene subtype in a downsampled WGS data set for 19 strains by both a blast and *in silico* PCR approach.

Trait	Number of strains with trait by real-time PCR	Number strains determined from WGS data with indicated trait at different sequence coverage levels									
		Original sequence coverage		40x coverage		30x coverage		20x coverage		10x coverage	
		Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR	Blast	<i>In silico</i> PCR
<i>stx1a</i>	2	3	2	2	2	2	2	2	2	2	2
<i>stx1c</i>	1	1	1	1	1	1	1	1	1	1	0
<i>stx1d</i>	2	2	2	2	2	2	2	2	2	2	1
<i>stx2a</i>	2	2	2	2	2	2	2	1	2	2	1
<i>stx2b</i>	3	2	3	2	3	2	3	3	3	3	2
<i>stx2c</i>	2	0	1	0	1	0	1	0	0	0	0
<i>stx2d</i>	3	3	3	3	3	3	3	3	2	3	2
<i>stx2e</i>	1	1	1	1	1	1	1	1	1	1	1
<i>stx2f</i>	7	7	7	7	7	7	7	7	7	7	6
<i>stx2g</i>	1	1	1	1	1	1	1	1	1	1	1

see Table S1 for isolate identification.

positive from the WGS using *in silico* PCR. One isolate was positive for *stx1a* using blast against the virulence gene database but was negative by both conventional testing using PCR and the *in silico* approach. Looking further into this discrepancy, using the blast based approach the gene was only a 82.3% length match compared to the reference allele and may indicate the gene was truncated so that the reverse primer used in the traditional or *in silico* PCR assay would not hybridize. Overall, *stx* gene subtype was correctly predicted in 89.5% of isolates at $\geq 10x$ coverage.

Validation of the Resistance Finder Tool in BioNumerics

The resistance finder tool in BioNumerics was evaluated against a set of 46 isolates where WGS and traditional antimicrobial susceptibility results were available. Several of the isolates tested, a total of 28 out of the 46, were pan-susceptible by both antimicrobial susceptibility testing and did not contain resistance genes by WGS (see Table 5). Of

the remaining STEC and *Shigella* that were found to be resistant by traditional antimicrobial susceptibility testing, the concordance for detecting genetic antimicrobial resistance determinants for ampicillin, azithromycin, chloramphenicol, sulfisoxazole, streptomycin, and trimethoprim/sulfamethoxazole was 100%. One isolate was found to contain tetracycline resistance genes that did not test as resistant by conventional testing. No genetic resistance determinants were detected for isolates resistant to nalidixic acid and ciprofloxacin using the ResFinder database. Through further genetic analysis, it was determined that these isolates were resistant via chromosomal mutations in the *gyrA* gene alone or in combination with mutations in the *parC* gene. These results are not unexpected as gene detection schemes can identify non-functional genes and do not detect mutational events. Taking these issues into account, there was 99.7% concordance between phenotypic susceptibility and antimicrobial resistance detection by WGS.

TABLE 5 | Comparison of phenotypic antimicrobial susceptibility testing results with resistance determinants identified from WGS data in 46 strains.

Antimicrobial	Phenotypic laboratory testing	Resistance determinant detection from WGS data
Ampicillin	12	12
Azithromycin	13	13
Chloramphenicol	1	1
Streptomycin	18	18
Sulfisoxazole	18	18
Nalidixic acid	9	0
Ciprofloxacin	6	0
Trimethoprim/Sulfamethoxazole	18	18
Tetracycline	17	18
No Resistance detected	28	28

Values indicate the number of strains identified with resistance to the indicated antimicrobial.

DISCUSSION

In this study, we demonstrated the utility and accuracy of a single software platform for combining workflows for quality assessment and reference characterization of STEC through WGS data. A single software program that can be used by non-bioinformaticians is a requirement for public health professionals to be able to infer phenotypic results from WGS data. Using publically available databases and *in silico* PCR tools developed as part of this study, we identified the same information (serotype, virulence genes, and resistance determinants) from the WGS data for 94.7% *E. coli* and *Shigella* isolates as was identified previously by conventional methods. Additionally, the limit of detection for these determinants was established through down sampling experiments allowing for better interpretation of negative results and understanding of sequence data quality needed for reference characterization from WGS.

Although, other recent studies have already shown the utility of reference characterization directly from sequence data generated from benchtop sequencers (Joensen et al., 2014, 2015; DebRoy et al., 2016), we present the combined quality assessment, serotyping, virulence profile, and resistance profile in one simple, high-throughput, and user-friendly analytical WGS workflow. These previous studies extensively validated their findings against those obtained by conventional methods, yet limited testing was done to identify the limit of detection for these tools and how best to interpret a negative result. Often in these studies, sequences were selected because they were of high quality and had high coverage, typically over 50x. While high sequence quality and coverage may be possible during routine testing periods, it is often difficult to achieve during outbreak response or when trying to reduce testing costs. When there is a need to increase isolate multiplexing per sequencing run to increase throughput and reduce costs, sequence coverage per isolate decreases. For this reason we attempted to determine the

limits for coverage to help determine the maximum number of isolates that could be sequenced at the same time. In our study, sequence coverage of ≥ 30 was enough to predict 100% of the H serogroups, for the O serogroups 93% of the serogroups were correctly predicted at a sequence coverage > 40 x. One O antigen, O153 (2 isolates), was not detected in genomes sequenced to over 100x coverage. Since other groups have shown that O153 genes *wzx* and *wzy* are 100% identical to the O178 genes, even though surprisingly these two serogroups are not cross-reactive using phenotypic testing, the current similarity-based WGS detection methods may not be able to distinguish these closely-related serogroups (Joensen et al., 2014, 2015; DebRoy et al., 2016).

Virulence gene detection performed more robustly than O and H antigen gene detection. The majority of virulence genes were detected at ≥ 20 x coverage in the WGS data using the genotyping plug-in. Since both serotype and virulence gene information is needed for STEC surveillance, to be able to consistently identify serotype from WGS in all isolates > 40 x coverage is recommended. Preliminary data (not presented) indicates that this coverage will also suffice for subtyping for outbreak investigations using whole genome multilocus sequence typing analysis.

To improve confidence in negative WGS results, additional *in-silico* PCR tools were developed to double check negative results from whole genome sequence data. This provided further confidence in virulence typing results. By using both WGS typing tools, all virulence genes were detected, by relying on either tool alone, blast algorithm or *in silico* PCR, important virulence genes would have been missed. Being able to accurately identify virulence genes and *stx* gene subtypes is important because certain virulence gene combinations are associated with higher risks for adverse events, e.g., HUS (Scheut et al., 2012). Other groups have also shown the robustness of determining *stx* gene subtypes from WGS from both O157 and non-O157 serogroups (Ashton et al., 2015; Chattaway et al., 2016).

For identification of resistance determinants, the ResFinder database produced highly concordant results with traditional phenotypic testing. Isolates possessing quinolone resistance mechanisms that were not identified by the ResFinder database, underscore the limitations of a gene-based detection approach. Supplementary *in silico* PCRs using conventional primer sets (Conrad et al., 1996; Bhattacharya et al., 2013) and subsequent sequence analysis or reference-mapping tools can be used in conjunction with the resistance finder tool, in order to detect mutational events conferring antimicrobial resistance. A recent study examining multi-drug resistant *E. coli* in the United States accurately predicted drug resistance with high specificity and sensitivity using a WGS approach that employed gene-based detection in conjunction with mutational analysis of the quinolone resistance-determining regions of the chromosome (Tyson et al., 2015). The present study confirms this high concordance between genetic and phenotypic testing for antimicrobial resistance, and also reveals the ability of this WGS-based approach to distinguish resistant and susceptible isolates for most drug classes.

From this study, it has been shown that quality assessment, serotyping, virulence and resistance profiling can be performed in one simple workflow. Additionally, the information that is extracted from WGS has more details than provided by conventional methods, e.g., by conventional methods we routinely detect only 5 virulence targets and 9 antimicrobial susceptibilities whereas over 100 virulence and resistance determinant genes are detected by WGS. Extracting this information from the whole genome sequence rather than using traditional identification techniques is highly cost-efficient: it is possible to save up to 180 US dollars on reagents alone per characterized isolate (assuming \$123 for WGS and \$304 for traditional typing workflow per isolate), which makes WGS both more rapid and less expensive for typing STEC. Additionally other groups (Joensen et al., 2014) have shown that the turnaround time for WGS is faster compared to conventional reference identification and subtyping workflows. We have also seen that the turnaround time from receipt of isolate in lab to WGS result can be 3–4 days while conventional methods take 1–3 weeks. While there is a high initial overhead cost of the sequencing instrument, WGS has the potential to streamline laboratory work into a unified workflow, reducing the need for multiple specialized personnel and instruments for various genotypic and phenotypic testing. Furthermore, by understanding the limit of detection by WGS for these different targets, there is more confidence that a negative result is an accurate prediction, though further work needs to be done to fine tune the serotype and resistance finder databases.

FUTURE WORK

For future work on the genotyping plug-in within BioNumerics, we plan to integrate more gene identification techniques from the whole genome sequence, such as reference mapping, to improve the ability to detect serogroups and resistance determinants. Lastly, although this validation was done for

reference characterization and a minimum coverage requirement was identified, similar tests need to be performed in terms of WGS analysis for outbreak detection. Currently we are validating a whole genome multilocus sequence typing (wgMLST) database for STEC.

AUTHOR CONTRIBUTIONS

RL, HC designed this work, RL, HC, JC, and HP contributed to the analysis and interpretation of the work. RL and HC drafted the manuscript and RL, HC, JC, HP, and NS reviewed the manuscript for intellectual content.

FUNDING

This work was made possible through support from the Advanced Molecular Detection (AMD) initiative at the Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

We are very grateful to Devon Stripling, Haley Martin, and Lisley Garcia-Toledo for help with traditional reference characterization and preparing isolates for sequencing as well as Ashley Sabol and Eija Trees in the PulseNet CDC team for providing some WGS data. We would also extend our thanks to Katrine Joensen and Flemming Schuetz for their excellent work on the SerotypeFinder and DTU for hosting the Finder databases. We would also like to thank PulseNet database managers Morgan Schroeder and Sung Im with assistance using BioNumerics v7.5.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00766>

REFERENCES

- Ashton, P. M., Perry, N., Ellis, R., Petrovska, L., Wain, J., Grant, K. A., et al. (2015). Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. *Peer J.* 3:e739. doi: 10.7717/peerj.739
- Bhattacharya, D., Thamizhmani, R., Bhattacharya, H., Sayi, D. S., Muruganandam, N., Roy, S., et al. (2013). Emergence of New Delhi metallo-beta-lactamase 1 (NDM-1) producing and multidrug resistant uropathogens causing urinary tract infections in Andaman Islands, India. *Microb. Drug Resist.* 19, 457–462. doi: 10.1089/mdr.2013.0070
- CDC (2013). *National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human Isolates Final Report*. Atlanta, Georgia: U.S. Department of Health and Human Services, CDC, 2015.
- Chattaway, M. A., Dallman, T. J., Gentle, A., Wright, M. J., Long, S. E., Ashton, P. M., et al. (2016). Whole genome sequencing for public health surveillance of Shiga Toxin-producing *Escherichia coli* other than serogroup O157. *Front. Microbiol.* 7:258. doi: 10.3389/fmicb.2016.00258
- Clinical and Laboratory Standards Institute, C. (2012). *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard*. Wayne, NJ: Pennsylvania.
- Conrad, S., Oethinger, M., Kaifel, K., Klotz, G., Marre, R., and Kern, W. V. (1996). gyrA mutations in high-level fluoroquinolone-resistant clinical isolates of *Escherichia coli*. *J. Antimicrob. Chemother.* 38, 443–455. doi: 10.1093/jac/38.3.443
- Cosentino, S., Voldby Larsen, M., Moller Aarestrup, F., and Lund, O. (2013). PathogenFinder—distinguishing friend from foe using bacterial whole genome sequence data. *PLoS ONE* 8:e77302. doi: 10.1371/journal.pone.0077302
- DebRoy, C., Fratamico, P. M., Yan, X., Baranzoni, G., Liu, Y., Needleman, D. S., et al. (2016). Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-typing. *PLoS ONE* 11:e0147434. doi: 10.1371/journal.pone.0147434
- Ewing, W. H. (1986). “The genus *Escherichia*,” in *Edward’s and Ewing’s Identification of Enterobacteriaceae, 4th Edn.*, ed W.H. Ewing (New York, NY: Elsevier Scientific Publishing Co., Inc.), 93–134.
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by

- use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Katz, L. S., Humphrey, J. C., Conley, A. B., Nelakuditi, V., Kislyuk, A. O., Agrawal, S., et al. (2011). Neisseria Base: a comparative genomics database for *Neisseria meningitidis*. *Database (Oxford)* 2011:bar035. doi: 10.1093/database/bar035
- Kleinheinz, K. A., Joensen, K. G., and Larsen, M. V. (2014). Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* 4:e27943. doi: 10.4161/bact.27943
- Paton, A. W., and Paton, J. C. (1998). Detection and characterization of Shiga toxinogenic *Escherichia coli* by using multiplex PCR assays for stx1, stx2, eaeA, enterohemorrhagic *E. coli* hlyA, rfbO111, and rfbO157. *J. Clin. Microbiol.* 36, 598–602.
- Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., et al. (2011). Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis.* 17, 7–15. doi: 10.3201/eid1701.P11101
- Scheutz, F., Teel, L. D., Beutin, L., Piérard, D., Buvens, G., Karch, H., et al. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12
- Tyson, G. H., Mcdermott, P. F., Li, C., Chen, Y., Tadesse, D. A., Mukherjee, S., et al. (2015). WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *J. Antimicrob. Chemother.* 70, 2763–2769. doi: 10.1093/jac/dkv186
- Zankari, E., Hasman, H., Kaas, R. S., Seyfarth, A. M., Agerso, Y., Lund, O., et al. (2013). Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J. Antimicrob. Chemother.* 68, 771–777. doi: 10.1093/jac/dks496

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

HP is affiliated as an employee (chief operations officer) with the following organization: Applied Maths NV, Keistraat 120, B-9830 Sint-Martens-Latem, Belgium.

Copyright © 2016 Lindsey, Pouseele, Chen, Strockbine and Carleton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Phylogenetic Analyses of *Shigella* and Enteroinvasive *Escherichia coli* for the Identification of Molecular Epidemiological Markers: Whole-Genome Comparative Analysis Does Not Support Distinct Genera Designation

Emily A. Pettengill¹, James B. Pettengill² and Rachel Binet^{1*}

OPEN ACCESS

Edited by:

Pina Fratamico,
United States Department
of Agriculture – Agricultural Research
Service, USA

Reviewed by:

Edward G. Dudley,
Penn State, USA
Adam Peritz,
United States Department
of Agriculture – Agricultural Research
Service, USA

*Correspondence:

Rachel Binet
rachel.binet@fda.hhs.gov

Specialty section:

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

Received: 13 November 2015

Accepted: 28 December 2015

Published: 19 January 2016

Citation:

Pettengill EA, Pettengill JB
and Binet R (2016) Phylogenetic
Analyses of *Shigella*
and Enteroinvasive *Escherichia coli*
for the Identification of Molecular
Epidemiological Markers:
Whole-Genome Comparative Analysis
Does Not Support Distinct Genera
Designation. *Front. Microbiol.* 6:1573.
doi: 10.3389/fmicb.2015.01573

¹ Division of Microbiology, Office of Regulatory Science, U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, College Park, MD, USA, ² Division of Public Health Informatics and Analytics, Office of Analytics and Outreach, U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, College Park, MD, USA

As a leading cause of bacterial dysentery, *Shigella* represents a significant threat to public health and food safety. Related, but often overlooked, enteroinvasive *Escherichia coli* (EIEC) can also cause dysentery. Current typing methods have limited ability to identify and differentiate between these pathogens despite the need for rapid and accurate identification of pathogens for clinical treatment and outbreak response. We present a comprehensive phylogeny of *Shigella* and EIEC using whole genome sequencing of 169 samples, constituting unparalleled strain diversity, and observe a lack of monophyly between *Shigella* and EIEC and among *Shigella* taxonomic groups. The evolutionary relationships in the phylogeny are supported by analyses of population structure and hierarchical clustering patterns of translated gene homolog abundance. Lastly, we identified a panel of 254 single nucleotide polymorphism (SNP) markers specific to each phylogenetic cluster for more accurate identification of *Shigella* and EIEC. Our findings show that *Shigella* and EIEC are not distinct evolutionary groups within the *E. coli* genus and, thus, EIEC as a group is not the ancestor to *Shigella*. The multiple analyses presented provide evidence for reconsidering the taxonomic placement of *Shigella*. The SNP markers offer more discriminatory power to molecular epidemiological typing methods involving these bacterial pathogens.

Keywords: *Shigella*, enteroinvasive *E. coli* (EIEC), phylogeny, whole genome sequencing, classification, epidemiological markers

INTRODUCTION

Shigella species are a leading cause of bacterial diarrhea (Walker et al., 2010). Worldwide, it is estimated that 164.7 million people are infected by *Shigella* annually (495,000 of those people in the United States) often through contaminated food and water (Scallan et al., 2011). Enteroinvasive *Escherichia coli* (EIEC), like *Shigella*, can also cause dysentery-like symptoms (Taylor et al., 1988).

Shigella and EIEC are, in essence, strict human pathogens, sharing similar pathogenic mechanisms but their evolutionary relationship on a genomic level has not been determined. Although, the close relationship between *Shigella* and *E. coli* has been acknowledged since 1898 (reviewed by Lan and Reeves, 2002), in the 1940s Ewing proposed classifying the four species in the new genus *Shigella* (*S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*) based on the antigen characteristics of those species (Edwards and Ewing, 1986). Since that time, numerous studies have indicated that the phylogenetic history does not support this current classification (Pupo et al., 2000; Lan and Reeves, 2002; Escobar-Páramo et al., 2003; Lan et al., 2004; Sahl et al., 2015).

Volunteer feeding studies have shown that whereas 10 to a few 100 *Shigella* cells were enough to cause illness in healthy adults, the infective dose for three different EIEC strains was more in the 10^8 range, justifying the need for clinical medicine to maintain two separate genera (DuPont et al., 1971; Mathewson et al., 1985). However, considering that most governmental health agencies do not currently require reporting EIEC infections, their impact on diarrheal disease and their genetic diversity is not well-understood. The recent involvement of EIEC O96:H19 as the source of outbreaks severely affecting healthy individuals in Italy, Great Britain and a case reported in Spain illustrates that EIEC can be a potential threat to public health and provides new motivation for improving our understanding of EIEC for rapid and accurate identification (Escher et al., 2014; Michelacci et al., 2015; Pettengill et al., 2015). This new motivation is reinforced by a long established need to understand the evolutionary relationships between *Shigella*, EIEC and non-invasive *E. coli* for improved detection and surveillance.

Traditional microbiology differentiates *Shigella* from *E. coli* based on their physiological and biochemical characteristics, with EIEC being more metabolically active than *Shigella* (Edwards and Ewing, 1986). Sero-agglutination assays are afterward generally performed for the differentiation of members of the genus *Shigella*, but cross-reactivity with certain EIEC serotypes have been observed (Liu et al., 2008). Developing nucleic acid-based detection methods combining higher discriminatory power with low limits of detection are ideal but rely on the availability of suitable markers based on a wide diversity of isolates for that organism (Zhao et al., 2014). Currently, most molecular assays for the diagnosis of *Shigella* rely mainly on targeting the large ~220-kbp invasive plasmid that is also shared by EIEC and, hence, cannot differentiate between the pathogens (Binet et al., 2014). Although, two recent studies proposed PCR assays to distinguish between *Shigella* species (Sahl et al., 2015) or between *Shigella* and EIEC (Pavlovic et al., 2011), the first study did not include any EIEC in their exclusivity panel and the second study included only 18 isolates of *Shigella* and 11 isolates of EIEC in their inclusivity panel.

In this study, we studied the evolutionary relationships among a wide diversity of strains that represent the *Shigella* genus and closely related EIEC. Comprehensive phylogenetic analyses were performed to determine if *Shigella* and EIEC are distinct evolutionary groups. Genome similarity was then investigated using a Bayesian clustering method that does not impose the bifurcating structure of phylogenetic analyses.

Samples were then hierarchically clustered based on differences in abundance of predicted protein homologs to determine functional genomic differences. Lastly, we identified single nucleotide polymorphisms (SNPs) that were diagnostic of different phylogenetic clades that could be used to type and/or discriminate among those lineages.

MATERIALS AND METHODS

Growth of Strains, DNA Isolation, and Genome Retrieval

Pure culture isolates for 33 *Shigella* and *E. coli* strains (**Supplementary Table S1**) were grown from frozen stocks on Trypticase Soy Agar plates and incubated overnight at 37°C. A minimum of three colonies were then inoculated into either *Shigella* Broth (if *Shigella* sp.; Center for Food Safety and Applied Nutrition, 2001) or Trypticase Soy Broth (if EIEC strains) for DNA extraction after overnight growth at 37°C. Genomic DNA was extracted using DNeasy® Blood and Tissue kits (QIAGEN, Valencia, CA, USA) according to manufacturers' instructions. An additional 80 genomes (**Supplementary Table S1**) were retrieved in June 2014 from the NCBI SRA database using the SRA Toolkit v. 2.3.5-2 in fastq format¹. Assembled genomes from Sahl et al. (2015) were retrieved from NCBI in February 2015.

Library Construction, Genome Sequencing, and Sequence Data

DNA was quantified using the Qubit® 2.0 Fluorometer and the Qubit® HS Assay kit (Life Technologies, Foster City, CA, USA). Samples were diluted to 0.2 ng/μl and stored at -20°C until library preparation. Libraries were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina®, San Diego, CA, USA). Sequencing reactions were performed with the MiSeq v2 chemistries with 250 bp paired-end read lengths and a 500-cycle cartridge and processed on a MiSeq platform (Illumina®, San Diego, CA, USA) to obtain data in fastq format. All the sequencing data generated for this project are available through bioproject accessions PRJNA273284 and PRJNA230969 at the National Center for Biotechnology Information (NCBI).

Quality Control, Trimming, and Genome Assembly

Reads were trimmed and low quality bases (Q-scores < 20) filtered using the DynamicTrim program in SolexaQA v. 2.2 (Cox et al., 2010). Trimmed reads were then assembled using SPAdes v. 3.1.1 (Bankevich et al., 2012) with default settings. To ensure that assemblies were of high quality (e.g., low number of contigs and adequate total length), we obtained assembly statistics using the program Quast (Gurevich et al., 2013; **Supplementary Table S3**). Using the *de novo* assemblies from SPAdes, SNP matrices were produced using the reference-free approach implemented in kSNP v2.0 (Gardner and Hall, 2013). For the kSNP analyses we

¹<http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=show&f=software&m=software&s=software>

used a *k*-mer value of 21, which was identified as the best fitting value based on the auxiliary script kChooser provided with that software.

Although, kSNP produces three matrices (composed of “all,” “majority,” and “core” SNPs), we focused on the core matrix as it is a more conservative method for identifying variant sites and better suited to remove recombination/horizontally transferred genomic elements from the analysis. The core matrix contains no missing data meaning there is a nucleotide state at each position in the alignment for all individuals. For kSNP analyses that included the *Salmonella* genomes, a total of 660,234 SNPs were identified and the number of core SNPs was 2,348. Analyses without *Salmonella* genomes had a total of 598,876 SNPs and 7,062 core SNPs. Of the core SNPs, 385 (16%) including *Salmonella* genomes and 1556 (22%) excluding *Salmonella* genomes were homoplastic (non-informative) SNPs. The proportions of homoplastic are lower than other kSNP analyses of *E. coli* genomes (37.6%; Gardner and Hall, 2013).

Serotyping

Shigella species are routinely serotyped with Statens Serum Institute species specific pool antisera (Cedarlane, Burlington, NC, USA) upon reception and by an in-house multiplex PCR assay (Binet, personal communication). Although the serotype is also confirmed with serotype specific Denka Seiken agglutinating sera (Thermo Fisher Scientific, Lenexa, KS, USA) on a case-by case basis, we did not confirm the identity of the nine *Shigella* isolates we sequenced in this study at the serotype level since they came from reputable bacterial collections, i.e., ATCC and CDC (Supplementary Table S1). All EIEC strains we sequenced were, however, conventionally serotyped with polyclonal O antigens from Statens Serum Institute (Cedarlane, Burlington, NC, USA) using a boiling method detailed by the manufacturer. For the additional genomes added to the study, in the absence of isolates, pertinent information was obtained directly from NCBI as provided at the time of submission or from associated publications when available (Holt et al., 2012; Escher et al., 2014; Sahl et al., 2015).

Phylogenetic Analysis and Sample Labeling Designations

Using the core matrix produced by kSNP, phylogenetic inference analysis was performed using GARLI (Genetic Algorithm for Rapid Likelihood Inference) v. 2.0.1019 under the GTR + I + Γ model and other default settings; trees were visualized with Figtree v. 1.3.1 (Zwickl, 2006; Rambaut and Drummond, 2009). To estimate the best topology based on the observed data, we ran 1000 replicate analyses and present the tree with the highest likelihood value. To estimate topological support for the different relationships, we ran 1000 bootstrap replicates that were then summarized using the SumTrees utility within the DendroPy package (Sukumaran and Holder, 2010). We chose not to remove homoplasious sites because bacterial phylogenetic topologies have been shown to be robust to the inclusion of such sites and removing them may in fact be detrimental to estimates of branch length (Hedge and Wilson, 2014).

Escherichia coli strains present in the phylogenetic tree are listed by the type of *E. coli*, the O antigen and H antigen (if known) followed by strain replicate number in parentheses. Other abbreviations found in the tree are: EIEC: enteroinvasive *E. coli*; EAEC: enteroaggregative *E. coli*; STEC: Shiga-toxin producing *E. coli*; ExPEC: extraintestinal pathogenic *E. coli*; EPEC: enteropathogenic *E. coli*; EHEC: enterohemorrhagic *E. coli* (Clements et al., 2012). *Shigella* strains are designated by genus and species, serotype (if known) followed by strain replicate number in parentheses. Abbreviations for *Shigella* species are as follows: SD: *S. dysenteriae*; SF: *S. flexneri*; SB: *S. boydii*, SS: *S. sonnei*.

Diagnostic SNP Detection

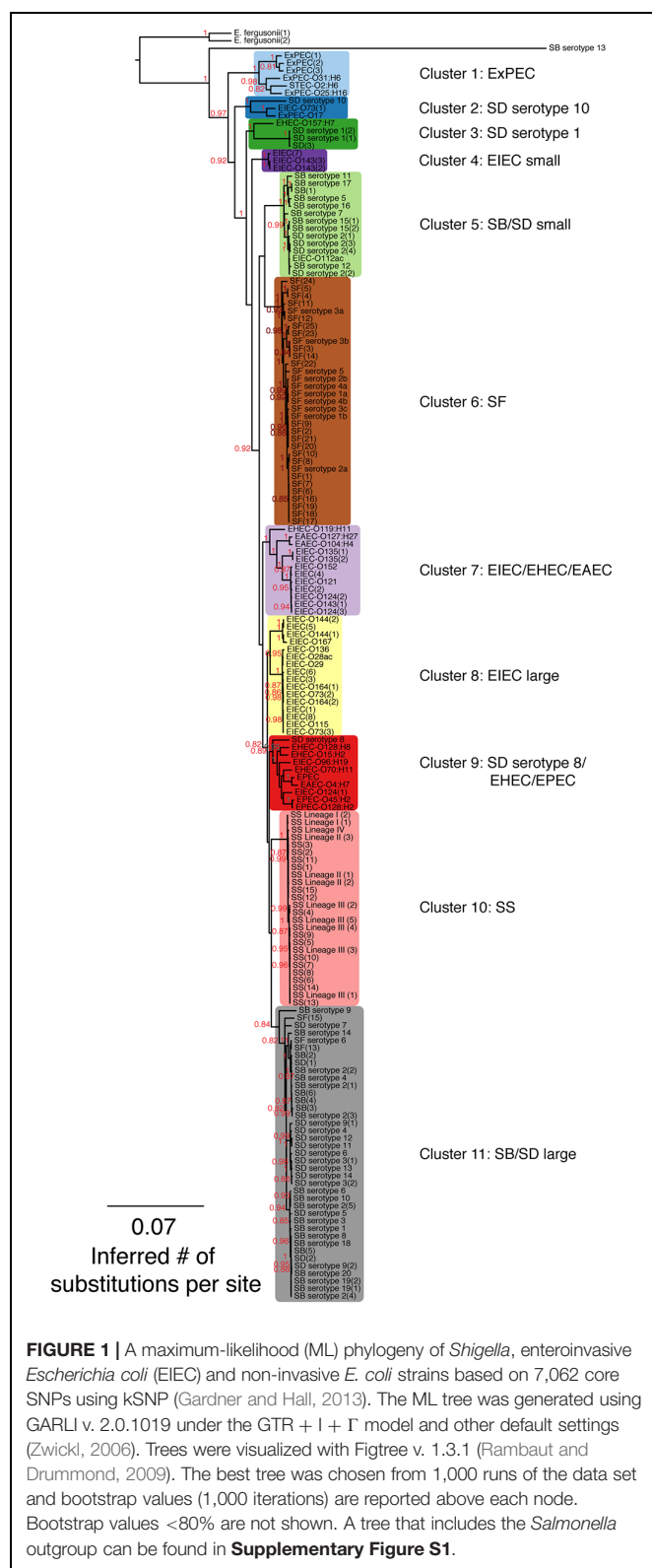
A separate kSNP analysis was performed without the two *Salmonella* outgroup samples to obtain a core SNP matrix for only *Shigella* and EIEC samples (described above). A custom python script was used with the core matrix to identify those SNPs that were specific to the groups from the SNP-based phylogeny. We define a diagnostic SNP as a position in the core matrix where the nucleotide state is the same among all members of a group and that state differs from all non-members. For each diagnostic SNP (Supplementary Table S2), we report the SNP nucleotide region of 21 bp (or *k*-mer), the diagnostic SNP state of that cluster, the position in relation to a reference genome (SD serotype 1, NCBI: CP000034), the name of the gene (if applicable), the product (if applicable), the functional Clusters of Orthologous Groups of proteins (COG) category and the reference genome locus tag (if applicable).

STRUCTURE Analyses

The STRUCTURE program performed model-based Bayesian clustering of genomes using the core SNP matrix without *Salmonella*, *E. fergusonii* or SB serotype 13 (related to *E. albertii*; Pritchard et al., 2000; Falush et al., 2003). Default parameters that consider admixture, were run for values of *k* between 2 through 11. The best fitting value of *k* identified by STRUCTURE HARVESTER based on changes in likelihood scores across the values of *k* as well as results from the value of *k* corresponding to the number of phylogenetic cluster (Evanno et al., 2005; Earl and vonHoldt, 2012). We ran ten replicate STRUCTURE runs for *k* = 2 to 11, each consisting of 6×10^4 generations, the first 10^4 served as the burn-in. Analyses were visualized using the DISTRUCT program (Rosenberg, 2004).

Genome Annotation, Homology Prediction, and Similarity Matrix

Genome annotation was performed with RAST v. 2.0 (ClassicRAST; Overbeek et al., 2013). Annotated genomes were used to predict the homology of predicted proteins using the GET_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013) program which uses a BLASTP bidirectional best hit approach with the following parameters: 75% amino acid sequence coverage, 1e-05 E-value and 60% sequence identity. This produced an abundance matrix of 3,777 predicted protein homologs that were identified in at least two genomes. Manhattan



distances were calculated from this matrix and clustered using the average linkage method with the hclust function in R Core Team (2014). Hierarchical clusters are colored to match the

phylogenetic clusters in **Figure 1** in a bar next to the heat map. To obtain bootstrap probabilities (BPs) for the dendrogram and assign approximately unbiased *p*-values (AU), the Pvcust program in R was used with 10,000 replicates and shown next to a heat map generated with ggplot2 (Suzuki and Shimodaira, 2006; Wickham, 2009; R Core Team, 2014).

Antibiotic Resistance-Related Annotation and Hierarchical Clustering

Using all genomes except those from Sahl et al. (2015) study, antibiotic resistance, the genes of antibiotic targets and biosynthesis genes were determined from a local BLASTN search using files available from the Comprehensive Antibiotic Resistance Database (downloaded in January 2015) with parameters set to an E-value of 1e-06 and 75% identity (**Supplementary Figure S4**; McArthur et al., 2013). The data were filtered to include genes that were present in at least two genomes. Hierarchical clustering, bootstrap support and approximately unbiased *p*-values were determined as described above.

Evaluation of Previously Described Molecular Assays for the Differentiation of *Shigella* and EIEC

Sahl et al. (2015) reported 11 primer pairs that were specific to their phylogenetic analysis of *Shigella* but they did not include EIEC strains in their analysis. Similarly, Pavlovic et al. (2011) reported that primers targeting the β -glucuronidase gene (*uidA*) and the lactose permease gene (*lacY*) could differentiate 18 isolates of *Shigella* from 11 isolates of EIEC. The primers sequence identities from those two studies were examined, *in silico*, using local BLAST searches against the 169 genomes in our analyses. In **Supplementary Figures S5** and **S6**, genomes for which the particular primer pair exhibited 95% or greater, and 92% or greater similarity, respectively, were shown in blue to predict PCR amplification. The figures were made using R Core Team (2014).

RESULTS

Phylogeny

One hundred and seventy-one genomes were selected to encompass a large selection of EIEC strains and represent the diversity of the *Shigella* genus. Genomes from 35 isolates were in-house sequenced draft genomes while 136 were available in public databases (**Supplementary Table S1**). We used 23 isolates of SD, including a minimum of 14 serotypes, 36 SF isolates, including at least six serotypes, 32 SB isolates, covering all 20 serotypes, 26 SS isolates, 32 EIEC isolates with 15 different serotypes, 18 isolates of non-invasive *E. coli* composed of 14 different serotypes, two isolates of *E. fergusonii*. The genomes of two *Salmonella* isolates were used for an outgroup (**Table 1**).

Single nucleotide polymorphisms found in every genome, defined as core SNPs, were used to generate SNP matrices. The kSNP v. 2.0 program (Gardner and Hall, 2013), which uses a *k*-mer based approach to identify variant sites across a set of genomes, generated SNP matrices consisting of 7,062

TABLE 1 | Number of bacterial isolates and serotypes.

Tree label	Description	Isolates	Serotypes
EIEC	Enteroinvasive <i>E. coli</i>	33	15
EAEC	Enteraggregative <i>E. coli</i>	3	3
STEC	Shiga-toxin producing <i>E. coli</i>	1	1
ExPEC	Extraintestinal <i>E. coli</i>	6	3
EPEC	Enteropathogenic <i>E. coli</i>	3	2
EHEC	Enterohemorrhagic <i>E. coli</i>	5	5
<i>E. fergusonii</i>	<i>E. fergusonii</i>	2	1
SD	<i>Shigella dysenteriae</i>	23	14
SF	<i>Shigella flexneri</i>	36	6
SB	<i>Shigella boydii</i>	32	20
SS	<i>Shigella sonnei</i>	26	1
<i>S. enterica</i>	<i>Salmonella enterica</i>	2	1
Total		171	72

or 2,348 core SNPs depending on whether the *Salmonella* outgroup was excluded (Figure 1) or included (Supplementary Figure S1). Subsequent phylogenetic reconstruction based on both SNP matrices resolved 11 groups that did not follow the taxonomic classification of the samples, thus implying that *Shigella*, EIEC, and non-invasive *E. coli* were polyphyletic (Figure 1; Supplementary Figure S1). With the exception of the EIEC large cluster, all clusters had adequate bootstrap support (greater than 0.83). The phylogeny shows that SD serotype 1, SD serotype 8, SD serotype 10, and SB serotype 13 do not cluster with any other *Shigella* serotypes (Figure 1). Clusters 1, 2, 3, 7, and 9 were composed of either EIEC or *Shigella* strains in combination with non-invasive *E. coli* strains, whereas clusters 4, 5, 6, 8, 10, and 11 contained only EIEC or *Shigella*. Clustering of SB and SD genomes suggests there are not distinct SB and SD lineages. Most SF genomes clustered together except those of SF serotype 6 that falls into cluster 11 with several serotypes of SB and SD. In the absence of actual isolates for SF(13) and SF(15) to conventionally determine their O-antigen type by sero-agglutination, we turned to molecular serotyping targeting the *wzx* and *wzy* genes involved in the assembly of the O-antigen. Gene alignments between SF(13) and SF(15) and *S. flexneri* serotype 6 *wzx* and *wzy* genes were 99% homolog (data not shown) and both strains identified as *E. coli* O147, which is nearly identical to *S. flexneri* type 6 (Liu et al., 2008), using the SerotypeFinder software (v. 1.1) accessible on the Center for Genomic Epidemiology server². For perspective on how many SNP differences are represented by the branch lengths, histograms of the pairwise distances of total SNP number between pairs of genomes can be found in Supplementary Figure S2.

Population Structure of SNP Clustering

Genome similarity was then investigated using a Bayesian clustering method that does not impose the bifurcating structure of phylogenetic analyses. The population structure of the samples was therefore examined using the Bayesian model-based program STRUCTURE v. 2.3.4. the core SNP matrix from

the kSNP program without *Salmonella* as input. The program assigns individuals to a fixed number of clusters (k) allowing for admixture (e.g., recombination, ancestral polymorphism, horizontal gene transfer). The program STRUCTURE Harvester was used to infer the optimal value of k that best fits the data (Evanno et al., 2005; Earl and vonHoldt, 2012), which was determined to be 6 (Figure 2A). We also chose a k value of 11 to represent the number of clusters in the phylogenetic analyses (Figures 1 and 2B). Both cluster schemes were similar to the phylogeny, particularly for SS, SF, ExPEC, and EIEC lineages and the two distinct SB/SD lineages (Figure 1). Genomes in clusters that include SD serotype 1, SD serotype 10, and SD serotype 8 shared core SNPs with genomes in the EIEC, ExPEC and very small proportions of SF and SS clusters (Figures 2A,B). When core SNPs from SF genomes were grouped into 11 genetic groups, the phylogeny topology was similar to that of the six groups with the exception of the SF genomes which appear to have two genetic backgrounds and these roughly correspond to the clustering observed in the phylogeny (Figure 1; Supplementary Figure S3).

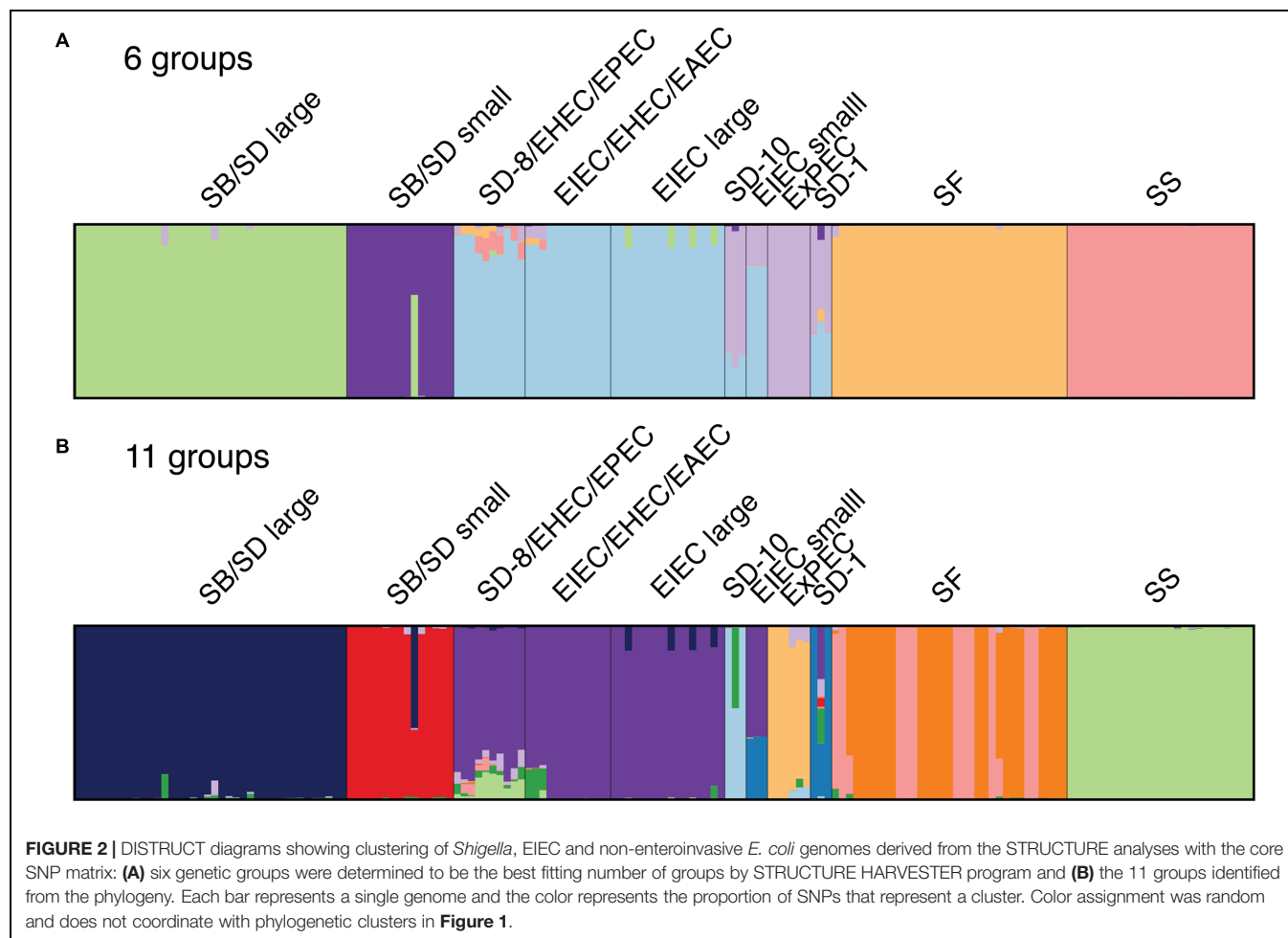
Clusters of Predicted Protein Homologs

The differences between the gene content of the genomes was then investigated based on the abundance of predicted protein homologs. After annotating all genomes with RAST (Overbeek et al., 2013), homologous translated genes were identified using the program GET_HOMOLOGUES which uses a BLASTP bidirectional best-hit approach (Contreras-Moreira and Vinuesa, 2013). While restricting our analyses to the genes that were shared between at least two individuals, we obtained a matrix composed of 3,777 genes and their abundances within each genome. The abundance matrix was hierarchically clustered with the average linkage method and Manhattan distances to identify differences in these profiles using the R package Pvcust (Suzuki and Shimodaira, 2006; R Core Team, 2014). Pvcust was also used to obtain statistical support for clusters based on both AU p -values and BP (Suzuki and Shimodaira, 2006). This showed that genomes from the phylogeny in the SS, SF, and SB/SD large clusters have significantly clustered translated gene abundance profiles with BP and AU of 100/100, 100/100, and 93/97, respectively (Figure 3). Hierarchical clustering of antibiotic resistance related genes shows patterns that are consistent with these studies and may indicate lineage specific selection in SS and some SD serotypes (Supplementary Figure S4).

Lineage-Specific SNP Identification and Evaluation of Previously Described Molecular Assays for the Differentiation of *Shigella* and EIEC

To identify lineage specific SNPs, we excluded the *Salmonella* outgroup to focus on differentiating among *Shigella* and EIEC lineages. From 7,062 core SNPs, we found 254 SNP positions that were diagnostic for each of the clusters (Supplementary Table S2). A description of the diagnostic SNPs by phylogenetic cluster is found in Table 2.

²<https://cge.cbs.dtu.dk/services/>



To illustrate the importance of using a genetically diverse set of genomes for the development of molecular epidemiological markers, we performed *in silico* analyses of primer sequence identities using BLAST searches for each primer against the full set (169) of genomes with a sequence identity of 95% (one base pair difference per primer) or higher for primers from (Sahl et al., 2015) or 92% and higher for primers from (Pavlovic et al., 2011; **Supplementary Figures S5 and S6**). We predict that these primers would not accurately distinguish between the phylogenetic groups determined by Sahl et al. (2015) or between *Shigella* and EIEC genomes, as suggested by Pavlovic et al. (2011; **Supplementary Figures S5 and S6**).

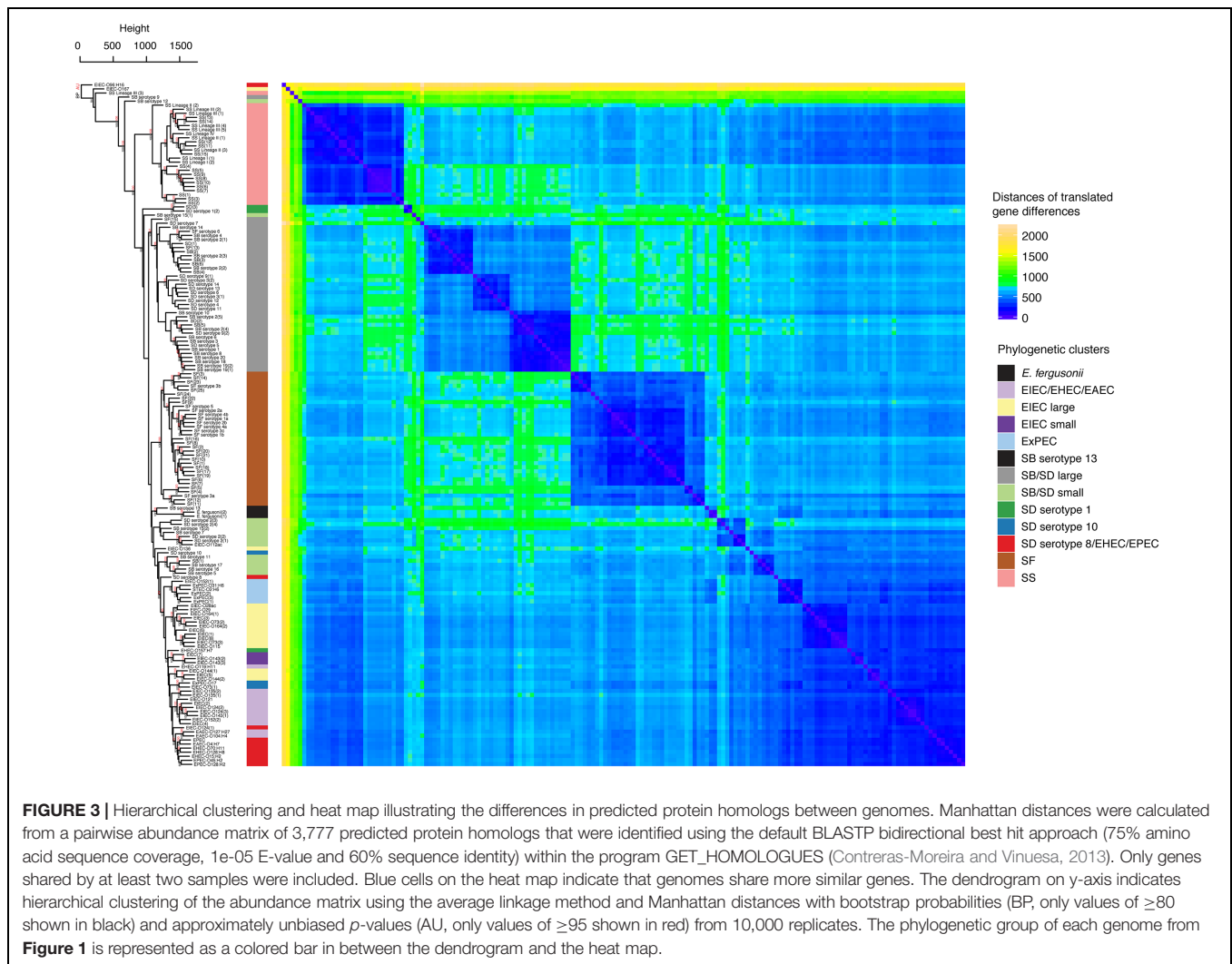
DISCUSSION

To the best of our knowledge, this study represents the most comprehensive phylogeny of *Shigella* and EIEC to date. Unlike previous studies exploring the molecular relationships between *E. coli* and *Shigella* (Pupo et al., 2000; Lan and Reeves, 2002; Escobar-Páramo et al., 2003, 2004; Lan et al., 2004; Touchon et al., 2009; Sims and Kim, 2011; Zhang and Lin, 2012; Gardner and Hall, 2013; Zuo et al., 2013; Sahl et al., 2014, 2015), we

used a large number and diversity of *Shigella* and EIEC genomes, including the recently discovered SB serotypes 19 and 20 and SD serotype 15, and performed genomic-scale phylogenetic analyses. The phylogeny together with the population structure analyses and the clustering of translated gene abundance profiles suggest that *Shigella* and EIEC evolved independently (**Figures 1–3**). Due to the polyphyly observed for EIEC, EIEC as a group cannot be considered as the ancestor to *Shigella* although some EIEC lineages may be the ancestor to *Shigella* (**Figure 1**). Interestingly, the phylogeny obtained is similar to the ones constructed using multi locus genotype data and other inference methods (i.e., neighbor-joining; Pupo et al., 2000; Lan and Reeves, 2002; Escobar-Páramo et al., 2003; Lan et al., 2004).

Incongruence between Phylogeny and Taxonomy

A few studies have concluded that *Shigella* arose from a single common ancestor (or monophyletically; Escobar-Páramo et al., 2003; Zuo et al., 2013). This conclusion likely comes from phylogenetic analyses conducted with a limited diversity of *Shigella* strains and serotypes and EIEC isolates. Analyses that include a broader diversity of strains support a hypothesis of multiple origins (Pupo et al., 2000; Lan and Reeves, 2002; Lan



et al., 2004; Sahl et al., 2015). Although many topological characteristics of our SNP-based phylogeny, such as the polyphyly of SB/SD, have been identified previously (Pupo et al., 2000; Lan and Reeves, 2002; Escobar-Páramo et al., 2003, 2004; Lan et al., 2004; Sahl et al., 2015), we clearly show that *Shigella* and EIEC genomes originated from multiple independent events. Similarly, the grouping of SF serotype 6 near SB serotypes 2, 4, and 14 indicates that, despite being called SF, they are part of the SB/SD large lineage (Pupo et al., 2000; Lan et al., 2004). As expected from previous studies that link SB serotype 13 to *E. albertii* (Pupo et al., 2000; Lan and Reeves, 2002; Hyma et al., 2005), our SB serotype 13 representative genome clusters outside of *E. coli*, EIEC and *Shigella* groups where it appears as the base of the phylogeny on an exceptionally long branch.

When considering EIEC specifically, our results are in agreement with those of Lan et al. (2004) where O124, O152, and O135 serotypes cluster together and O136, O28ac, O164, and O29 cluster together. Similarly, we observed, that EIEC serotype O112ac clustered near SB serotype 12 and SD serotype 2, and

identified only five core SNP differences between EIEC serotype O112ac and SD serotype 2(2) (Figures 1–3).

One topological difference between our phylogeny and previous phylogenies is the clustering of SB serotype 12. In our phylogenetic analyses (Figure 1), SB serotype 12 clusters in the SB/SD small cluster as opposed to clustering with SF strains in trees constructed by Pupo et al. (2000) and Lan et al. (2004). Our kSNP analyses reveal that there are only eight core SNP differences between SB serotype 12 and SD serotype 2(1). However, clustering of the translated gene abundance matrix shows that SB serotype 12 clusters by itself, away from any isolates it clusters near in the phylogeny (Figure 3). This suggests that SB serotype 12 may have a unique genetic history requiring additional analyses.

Given that we did not remove homoplastic SNPs based on the phylogenetic results, we can infer the degree of admixture (perhaps due to recombination) among the samples based on the STRUCTURE results. In general, both the clustering at $k = 6$ and 11 show only a few samples to have SNP profiles that suggest admixture with other distinct groups. Also from the

TABLE 2 | Phylogenetic group name (from **Figure 1**), number of individuals within each group (*N*) and the number of diagnostic SNPs (*D_{snps}*).

Group	<i>N</i>	<i>D_{snps}</i>
EIEC/EHEC/EAEC	12	6
EIEC large	16	0
EIEC small	3	31
ExPEC	6	71
SB/SD large	38	7
SB/SD small	15	21
SD serotype 1	3	1
SD serotype 10	3	37
SD serotype 8/EHEC/EPEC	10	1
SF	33	34
SS	26	45
Total	165	254

STRUCTURE analyses, we see that hybrid strains within SB, SD, and SS lineages may be rare. An exception is SB serotype 9 (**Figure 3**) and, similar to SB serotype 12 discussed above, the hierarchical clustering of the translated gene abundance matrix shows it clustering distantly from strains it clusters near in the phylogeny. It would be interesting to further investigate a range of SB serotype 9 isolates to determine if this pattern is common and represents a transitional strain.

While we did not specifically investigate the evolutionary history of the invasion plasmid, our data do not support the hypothesis proposed by Escobar-Páramo et al. (2003) that the invasion plasmid was transferred before the evolution of *Shigella* and EIEC lineages. Our phylogeny and the DISTRUCT diagram (**Figures 1** and **2**) suggest that EIEC cluster with non-invasive *E. coli* genomes that do not possess the invasion plasmid implying that the transfer of the invasion plasmid did not precede a monophyletic evolution of *Shigella* and EIEC.

Importance of Sampling Diverse Genetic Lineages

Our study underscores the importance of including a diverse collection of *Shigella* and EIEC genomes into phylogenetic studies that examine *Shigella*, as we were able to make a number of novel findings with high confidence. For example, EIEC strains appear to have a greater genetic diversity than previously believed, with EIEC strains clustering near non-invasive *E. coli* strains (**Figure 1**). For this reason, the inclusion of a range of EIEC strains for developing diagnostic tools is essential for accurate and clinically relevant identification, as well as for outbreak detection. When genetic diversity is not a component of investigations for diagnostic purposes, markers may not be useful. One example is a recent study that presents diagnostic markers for PCR detection of *Shigella* (Sahl et al., 2015), yet the primers for these markers do not discriminate between *Shigella* and EIEC when a larger genetic diversity is considered (**Supplementary Figure S5**). While another study included 11 EIEC strains (Pavlovic et al., 2011), their primers and probes cannot accurately distinguish between *Shigella* and EIEC (**Supplementary Figure S6**).

Single nucleotide polymorphisms markers are a useful genotyping/molecular epidemiological typing method because they are considered relatively genetically stable and not likely to change, to such a degree that classification tools are built based on SNPs (Larkeryd et al., 2014). Another asset is that a nucleotide should always be present at the SNP position, reducing the number of false negatives from presence/absence-type gene markers. SNP detection methods are also considered excellent for their discriminatory power, reproducibility and ability to be used in a high-throughput capacity (Hallin et al., 2012). With these advantages in mind, we identified multiple SNPs for the phylogenetic groups (except EIEC large), which offer researchers multiple opportunities for optimizing primer design and confirming positive results. Our inability to identify diagnostic SNP markers for the EIEC large cluster suggests that a greater diversity of EIEC isolates would be needed for markers (**Supplementary Table S3**). The lower bootstrap support (0.61) for the EIEC large cluster (**Figure 1**) is consistent with a need for additional genomes with greater genetic diversity.

Analyses looking for the presence/absence of core genes that were specific to each cluster yielded no such genes. This is in agreement with another study that did not identify *Shigella*-specific genes that were distinct from *E. coli* using orthologous genes from pan-genomes (Gordienko et al., 2013). As the authors and our data suggest, phylogenetic evidence points toward *Shigella* belonging to the *E. coli* genus and thus these groups are likely sharing the same pool of genes. In summary, the polyphyletic nature of the *Shigella* and *E. coli* groups and putative taxonomy makes the strategy of identifying specific genes to these groups difficult.

The clustering based on the abundance matrix of translated genes is not strictly congruent with the topology inferred from the phylogenetic analyses using the SNP data (**Figures 1** and **3**). However, most of the incongruence is among clades rather than the membership of individuals to specific clades. For example, all but one of the individuals belonging to the SF, SB/SD large, and SS clusters are not found grouped together in the trees based on the SNP and gene abundance data but the relationships among those clades does differ (**Figures 1** and **3**). Overall, we find support that gene content/abundance carries a similar evolutionary signal as that contained in SNPs. For example, there is an appreciable amount of resolution and fidelity to the relationships depicted in the phylogeny using the hierarchically clustered distance matrix of predicted protein homologs for clusters of genomes from SF, SS, SB/SD large clusters (**Figure 3**). These clusters have significant AU values and strong bootstrap support. However, differences do exist between the methods, which may be the result of unresolved basal relationships and/or unique isolate outliers (such as EIEC O96:H16, SB serotypes 9 and 12). It is also possible that the gene abundance analyses are capturing a stronger signal from recombination and mobile elements than would be present in the core SNP matrix. A similar incongruence was observed in a very limited number of *Shigella* and *E. coli* genomes between phylogenies based on core SNPs and using BLAST derived coding sequences (CDSs; Sahl et al., 2014). Some degree of incongruence is to be expected due to gene histories being linked but different from species histories (Szöllosi et al., 2015). For example, studies

of SS and SD provide evidence that these lineages are undergoing selection for drug and multidrug resistance and we also observed a pattern of clustering of antibiotic resistance-related genes that are linked to phylogeny but also may have individual gene histories (**Supplementary Figure S4**; Holt et al., 2012; Rohmer et al., 2014).

CONCLUSION

There is a growing acknowledgment that microbial taxonomy should be based on a more comprehensive and exhaustive survey of genomes (Rosselló-Móra and Amann, 2015; Thompson et al., 2015). Current problematic taxonomic designations are common throughout microbial taxonomy (Rosselló-Móra and Amann, 2015; Thompson et al., 2015 and references within). In the case of *Shigella*, genomic evidence supporting the change of taxonomic designations is well-established (Pupo et al., 2000; Lan and Reeves, 2002; Escobar-Páramo et al., 2003, 2004; Lan et al., 2004; Gardner and Hall, 2013; Sahl et al., 2014, 2015). Based on these studies and the analyses conducted herein, there is a large body of evidence that the *Shigella* genus should be moved back within the species *E. coli*. Furthermore, we suggest that *Shigella* should be classified as EIEC and the serotypes renamed using the common O antigen naming. *Shigella* serotypes are based upon O antigens many of which are identical or nearly identical to existing *E. coli* O antigens (with the exception of *S. sonnei*; Liu et al., 2008). The existence of two separate nomenclatures is redundant and confusing. We are repeating a long established call to reduce confusion and promote the understanding of accurate evolutionary relationships of *Shigella* and *E. coli* (Lan and Reeves, 2002; Chaudhuri and Henderson, 2012). While we believe that taxonomic designations that more accurately reflect genetic relationships can improve outbreak characterization and communication in the long-term, taxonomic revisions are difficult and some may consider that revisions pose risks for public health in the more immediate time frame. We support the growing recognition of the value behind systematic species or genome similarity assignments for all players involved in real-time epidemics (Marakeby et al., 2014; Varghese et al., 2015; Weisberg et al., 2015). In the absence of universal genome-based classification and naming systems, our results provide support for reconsidering the current taxonomic placement and naming of *Shigella* species.

AUTHOR CONTRIBUTIONS

EP generated sequence data. EP and JP performed analyses. EP, JP, and RB interpreted results and wrote the manuscript. RB conceived the project. All authors read and approved the final manuscript.

FUNDING

EP was funded in part by an appointment with the Research Participation Program at the Center for Food Safety and Applied

Nutrition administered by Oak Ridge Institute for Science and Education through and interagency agreement between the U.S. Department of Energy and The U.S. Food and Drug Administration.

ACKNOWLEDGMENTS

We would like to thank John Miller and Yan Luo for assistance with Python scripting and analyses, and Tom Hammack and Eric Brown for their support and critical reading of the manuscript. The authors would also like to thank Rosangela Tozzoli and Stefano Morabito for providing the raw data for EIEC O96:H19 before public release of the genome as well as Anthony Maurelli and the STEC center for providing isolates.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01573>

FIGURE S1 | A maximum-likelihood (ML) phylogeny of *Shigella*, enteroinvasive *E. coli* (EIEC), non-invasive *E. coli* strains and *Salmonella* outgroup based on 2,348 SNPs present in all genomes using the kSNP program (Gardner and Hall, 2013). The ML tree was generated using GARLI v. 2.0.1019 (Zwickl, 2006) under the GTR + I + Γ model and other default settings. Trees were visualized with Figtree v. 1.3 (Rambaut and Drummond, 2009). The best tree was chosen from 100 runs of the data set and bootstrap values (1,000 iterations) are reported above each node. Bootstrap values <80% were not shown.

FIGURE S2 | Histograms of the pairwise distances of core SNP differences between genome pairs for the SNP-based phylogenies (A) without the *Salmonella* outgroup and (B) with the *Salmonella* outgroup.

FIGURE S3 | Reordered STRUCTURE results for *S. flexneri* genomes from analyses performed with 11 SNP groups (right) corresponding to the phylogenetic cluster in **Figure 1** (left).

FIGURE S4 | Hierarchical clustering of antibiotic resistance related genes. Red values on dendrogram represent unbiased *p*-values determined by Pvcust package in R. The dendrogram was generated using the correlation distance method and the average linkage method.

FIGURE S5 | BLAST alignment of primers, described by Sahl et al. as specific for *Shigella* phylogenetic groups (Sahl et al., 2015), with genomes used in this study. A blue cell for a particular genome indicates that both primers of the pair aligned to 95% or greater sequence identity and should therefore hybridize to yield a PCR product. The phylogenetic group designation assigned by Sahl et al. is noted next to the cluster designations we observed with these genomes.

FIGURE S6 | *In silico* alignment of primer-probe sets described by Pavlovic et al. (2011) with genomes used in this study using BLAST. The *lacY* set was supposed to differentiate between *Shigella* (absent) and EIEC (present), while the *uidA* set was intended to be a positive control (present in both). BLAST identities of 92% or higher are shown with blue cells. Although PCR products are expected from a particular genome if both cells corresponding to the forward and reverse primers are highlighted in blue, the real-time PCR assay (Pavlovic et al., 2011) also require the respective probe to hybridize efficiently and therefore the respective cell to be highlighted in blue in the figure.

TABLE S1 | Strain information includes NCBI identifier (SRA#), Tree label/Strain designation, genus and species with serotype, O or H antigens, additional strain identifiers and reference for source of genomes.

TABLE S2 | Full list of diagnostic SNPs for *Shigella* and EIEC phylogenetic clusters. Includes phylogenetic cluster name, 21 bp sequence of region containing diagnostic SNP with ambiguous SNP state represented by “.”, diagnostic SNP state of cluster, position in the NCBI annotated reference genome (SD serotype 1, CP000034), gene name (“NA” if intergenic), functional gene product (“NA” if intergenic), COG identifier and reference genome (CP000034) locus tag.

TABLE S3 | Assembly statistics and genome metrics calculated by the Quast program. Includes Tree label/Strain designation, NCBI SRA accession number, number of contigs greater or equal to 1,000 bp (# contigs ($\geq 1,000$ bp)), number of contigs greater or equal to 0 bp ((# contigs (≥ 0 bp)), total length of contigs greater or equal to 1,000 bp (Total length ($\geq 1,000$ bp)), total length of contigs greater or equal to 0 bp (Total length (≥ 0 bp)), number of contigs, largest contig (bp), total length of all contigs, percent GC content and number of N's per 100 kbp.

REFERENCES

- 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
- Binet, R., Deer, D. M., and Uhlfelder, S. J. (2014). Rapid detection of *Shigella* and enteroinvasive *Escherichia coli* in produce enrichments by a conventional multiplex PCR assay. *Food Microbiol.* 40, 48–54. doi: 10.1016/j.fm.2013.12.001
- Center for Food Safety and Applied Nutrition (2001). *Laboratory Methods - Bacteriological Analytical Manual (BAM)*. Available at: <http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm2006949.htm>
- Chaudhuri, R. R., and Henderson, I. R. (2012). Infection, genetics and evolution. *Infect. Genet. Evol.* 12, 214–226. doi: 10.1016/j.meegid.2012.01.005
- Clements, A., Young, J. C., Constantinou, N., and Frankel, G. (2012). Infection strategies of enteric pathogenic *Escherichia coli*. *Gut Microbes* 3, 71–87. doi: 10.4161/gmic.19182
- Contreras-Moreira, B., and Vinuesa, P. (2013). GET_homologues, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13
- Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11:485. doi: 10.1186/1471-2105-11-485
- DuPont, H. L., Formal, S. B., Hornick, R. B., Snyder, M. J., Libonati, J. P., Sheahan, D. G., et al. (1971). Pathogenesis of *Escherichia coli* diarrhea. *N. Engl. J. Med.* 285, 1–9. doi: 10.1056/NEJM197107012850101
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Edwards, P. R., and Ewing, W. H. (1986). *Edwards and Ewing's Identification of Enterobacteriaceae*. Amsterdam: Elsevier Publishing Company.
- Escher, M., Scavia, G., Morabito, S., Tozzoli, R., Maugliana, A., Cantoni, S., et al. (2014). A severe foodborne outbreak of diarrhoea linked to a canteen in Italy caused by enteroinvasive *Escherichia coli*, an uncommon agent. *Epidemiol. Infect.* 142, 2559–2566. doi: 10.1017/S0950268814000181
- Escobar-Páramo, P., Clermont, O., Blanc-Potard, A. -B., Bui, H., Le Bouguénec, C., and Denamur, E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol. Biol. Evol.* 21, 1085–1094. doi: 10.1093/molbev/msh118
- Escobar-Páramo, P., Giudicelli, C., Parsot, C., and Denamur, E. (2003). The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J. Mol. Evol.* 57, 140–148. doi: 10.1007/s00239-003-2460-3
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.
- Gardner, S. N., and Hall, B. G. (2013). When whole-genome alignments just won't work: ksnpp v2 software for alignment-free snp discovery and phylogenetics of hundreds of microbial genomes. *PLoS ONE* 8:e81760. doi: 10.1371/journal.pone.0081760
- Gordienko, E. N., Kazanov, M. D., and Gelfand, M. S. (2013). Evolution of pan-genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica*. *J. Bacteriol.* 195, 2786–2792. doi: 10.1128/JB.02285-12
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hallin, M., Deplano, A., and Struelens, M. J. (2012). “Molecular typing of bacterial pathogens: a tool for the epidemiological study and control of infectious diseases,” in *New Frontiers of Molecular Epidemiology of Infectious Diseases*, eds S. Morand, F. Beaudeau, J. Cabaret (Dordrecht: Springer), 9–25.
- Hedge, J., and Wilson, D. J. (2014). Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* 5:e2158–e2164. doi: 10.1128/mBio.02158-14
- Holt, K. E., Baker, S., Weill, F.-X., Holmes, E. C., Kitchen, A., Yu, J., et al. (2012). *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat. Genet.* 44, 1056–1059. doi: 10.1038/ng.2369
- Hyma, K. E., Lacher, D. W., Nelson, A. M., Bumbaugh, A. C., Janda, J. M., Strockbine, N. A., et al. (2005). Evolutionary genetics of a new pathogenic *Escherichia* Species: *Escherichia albertii* and related *Shigella boydii* Strains. *J. Bacteriol.* 187, 619–628. doi: 10.1128/JB.187.2.619-628.2005
- Lan, R., Alles, M. C., Donohoe, K., Martinez, M. B., and Reeves, P. R. (2004). Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infect. Immun.* 72, 5080–5088. doi: 10.1128/IAI.72.9.5080-5088.2004
- Lan, R., and Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect.* 4, 1125–1132. doi: 10.1016/S1286-4579(02)01637-4
- Larkeryd, A., Myrtennas, K., Karlsson, E., Dwibedi, C. K., Forsman, M., Larsson, P., et al. (2014). CanSNPer: a hierarchical genotype classifier of clonal pathogens. *Bioinformatics* 30, 1762–1764. doi: 10.1093/bioinformatics/btu113
- Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., et al. (2008). Structure and genetics of *Shigella* O antigens. *FEMS Microbiol. Rev.* 32, 627–653. doi: 10.1111/j.1574-6976.2008.00114.x
- Marakeby, H., Badr, E., Torkey, H., Song, Y., Leman, S., Monteil, C. L., et al. (2014). A system to automatically classify and name any individual genome-sequenced organism independently of current biological classification and nomenclature. *PLoS ONE* 9:e89142. doi: 10.1371/journal.pone.0089142
- Mathewson, J. J., Johnson, P. C., DuPont, H. L., Morgan, D. R., Thornton, S. A., Wood, L. V., et al. (1985). A newly recognized cause of travelers' diarrhea: enteroadherent *Escherichia coli*. *J. Infect. Dis.* 151, 471–475. doi: 10.1093/infdis/151.3.471
- McArthur, A. G., Wagglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13
- Michelacci, V., Prosseda, G., Maugliani, A., Tozzoli, R., Sanchez, S., Herrera-León, S., et al. (2015). Characterisation of an emergent clone of enteroinvasive *Escherichia coli* circulating in Europe. *Clin. Microbiol. Infect.* doi: 10.1016/j.cmi.2015.10.025 [Epub ahead of print].
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2013). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226
- Pavlovic, M., Luze, A., Konrad, R., Berger, A., Sing, A., Busch, U., et al. (2011). Development of a duplex real-time PCR for differentiation between *E. coli* and *Shigella* spp. *J. Appl. Microbiol.* 110, 1245–1251. doi: 10.1111/j.1365-2672.2011.04973.x
- Pettengill, E. A., Hoffmann, M., Binet, R., Roberts, R. J., Payne, J., Allard, M., et al. (2015). Complete genome sequence of enteroinvasive *Escherichia coli* O96:H19 associated with a severe foodborne outbreak. *Genome Announc.* 3:e883–e885. doi: 10.1128/genomeA.00883-15

- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Pupo, G. M., Lan, R., and Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10567–10572. doi: 10.1073/pnas.180094797
- Rambaut, A., and Drummond, A. (2009). *FigTree v1. 3.1: Tree Figure Drawing Tool*. Available at: <http://tree.bio.ed.ac.uk/software/figtree>
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing Available at: <http://www.R-project.org>
- Rohmer, L., Jacobs, M. A., Brittnacher, M. J., Fong, C., Hayden, H. S., Hocquet, D., et al. (2014). Genomic analysis of the emergence of 20th century epidemic dysentery. *BMC Genomics* 15:355. doi: 10.1186/1471-2164-15-355
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol.* 4, 137–138.
- Rosselló-Móra, R., and Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* 38, 209–216. doi: 10.1016/j.syapm.2015.02.001
- Sahl, J. W., Caporaso, J. G., Rasko, D. A., and Keim, P. (2014). The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2:e332. doi: 10.7717/peerj.332
- Sahl, J. W., Morris, C. R., Emberger, J., Fraser, C. M., Ochieng, J. B., Juma, J., et al. (2015). Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J. Clin. Microbiol.* 53, 951–960. doi: 10.1128/JCM.03527-14
- Scallan, E., Hoekstra, R. M., and Angulo, F. J. (2011). Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis.* 17, 7–15. doi: 10.3201/eid1701.091101p1
- Sims, G. E., and Kim, S. H. (2011). Whole-genome phylogeny of *Escherichia coli/Shigella* group by feature frequency profiles (FFPs). *Proc. Natl. Acad. Sci. U.S.A.* 106, 8329–8334.
- Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571. doi: 10.1093/bioinformatics/btq228
- Suzuki, R., and Shimodaira, H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542. doi: 10.1093/bioinformatics/btl117
- Szöllosi, G. J., Tannier, E., Daubin, V., and Boussau, B. (2015). The inference of gene trees with species trees. *Syst. Biol.* 64, e42–e62. doi: 10.1093/sysbio/syu048
- Taylor, D. N., Echeverria, P., Sethabutr, O., Pitarangsi, C., Leksomboon, U., Blacklow, N. R., et al. (1988). Clinical and microbiologic features of *Shigella* and enteroinvasive *Escherichia coli* infections detected by DNA hybridization. *J. Clin. Microbiol.* 26, 1362–1366.
- Thompson, C. C., Amaral, G. R., Campeão, M., Edwards, R. A., Polz, M. F., Dutilh, B. E., et al. (2015). Microbial taxonomy in the post-genomic era: rebuilding from scratch? *Arch. Microbiol.* 197, 359–370. doi: 10.1007/s00203-014-1071-2
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5:e1000344. doi: 10.1371/journal.pgen.1000344
- 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. doi: 10.1093/nar/gkv657
- Walker, C. L. F., Sack, D., and Black, R. E. (2010). Etiology of Diarrhea in Older Children, adolescents and adults: a systematic review. *PLoS Negl. Trop. Dis.* 4:e768. doi: 10.1371/journal.pntd.0000768
- Weisberg, A. J., Elmarakeby, H. A., Heath, L. S., and Vinatzer, B. A. (2015). Similarity-based codes sequentially assigned to ebolavirus genomes are informative of species membership, associated outbreaks, and transmission chains. *Open Forum Infect. Dis.* 2:ofv024. doi: 10.1093/ofid/ofv024
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.
- Zhang, Y., and Lin, K. (2012). A phylogenomic analysis of *Escherichia coli/Shigella* group: implications of genomic features associated with pathogenicity and ecological adaptation. *BMC Evol. Biol.* 12:1. doi: 10.1186/1471-2148-12-174
- Zhao, X., Lin, C. -W., Wang, J., and Oh, D. H. (2014). Advances in rapid detection methods for foodborne pathogens. *J. Microbiol. Biotechnol.* 24, 297–312. doi: 10.4014/jmb.1310.10013
- Zuo, G., Xu, Z., and Hao, B. (2013). *Shigella* Strains are not clones of *Escherichia coli* but sister species in the genus *Escherichia*. *Genomics Proteom. Bioinform.* 11, 61–65. doi: 10.1016/j.gpb.2012.11.002
- Zwickl, D. J. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum-likelihood criterion*. Ph.D. thesis, University of Texas at Austin.

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

Copyright © 2016 Pettengill, Pettengill and Binet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Whole Genome Sequencing for Public Health Surveillance of Shiga Toxin-Producing *Escherichia coli* Other than Serogroup O157

Marie A. Chattaway, Timothy J. Dallman, Amy Gentle, Michael J. Wright, Sophie E. Long, Philip M. Ashton, Neil T. Perry and Claire Jenkins *

Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Remy Froissart,
Centre National de la Recherche
Scientifique, France
Seamus Fanning,
University College Dublin, Ireland

*Correspondence:

Claire Jenkins
claire.jenkins@phe.gov.uk

Specialty section:

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

Received: 16 January 2016

Accepted: 16 February 2016

Published: 03 March 2016

Citation:

Chattaway MA, Dallman TJ, Gentle A,
Wright MJ, Long SE, Ashton PM,
Perry NT and Jenkins C (2016) Whole
Genome Sequencing for Public Health
Surveillance of Shiga Toxin-Producing
Escherichia coli Other than Serogroup
O157. *Front. Microbiol.* 7:258.
doi: 10.3389/fmicb.2016.00258

Shiga toxin-producing *Escherichia coli* (STEC) are considered to be a significant threat to public health due to the severity of gastrointestinal symptoms associated with human infection. In England STEC O157 is the most commonly detected STEC serogroup, however, the implementation of PCR at local hospital laboratories has resulted in an increase in the detection of non-O157 STEC. The aim of this study was to evaluate the use of whole genome sequencing (WGS) for routine public health surveillance of non-O157 STEC by comparing this approach to phenotypic serotyping and PCR for subtyping the stx-encoding genes. Of the 102 isolates where phenotypic and genotypic serotyping could be compared, 98 gave fully concordant results. The most common non-O157 STEC serogroups detected were O146 (22) and O26 (18). All but one of the 38 isolates that could not be phenotypically serotyped (designated O unidentifiable or O rough) were serotyped using the WGS data. Of the 73 isolates where a flagella type was available by traditional phenotypic typing, all results matched the H-type derived from the WGS data. Of the 140 sequenced non-O157 isolates, 52 (37.1%) harboured *stx1* only, 42 (30.0%) had *stx2* only, 46 (32.9%) carried *stx1* and *stx2*. Of these, stx subtyping PCR results were available for 131 isolates and 121 of these had concordant results with the stx subtype derived from the WGS data. Of the 10 discordant results, non-specific primer binding during PCR amplification, due to the similarity of the *stx2* subtype gene sequences was the most likely cause. The results of this study showed WGS provided a reliable and robust one-step process for characterization of STEC. Deriving the full serotype from WGS data in real time has enabled us to report a higher level of strain discrimination while stx subtyping provides data on the pathogenic potential of each isolate, enabling us to predict clinical outcome of each case and to monitor the emergence of hyper-virulent strains.

Keywords: whole genome sequencing, Shiga Toxin-producing *Escherichia coli*, serotyping, stx subtyping

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are considered to be a significant threat to public health due to the severity of gastrointestinal symptoms associated with human infection and the risk of cases developing Haemolytic Uraemic Syndrome (HUS; Byrne et al., 2015). STEC are zoonotic; transmission occurs by direct contact with animals or their environment, or by consumption of

contaminated food or water (Byrne et al., 2014). The infectious dose is low (<10 organisms) and person-to-person spread is common, particularly in nursery school settings and in households with young children (Byrne et al., 2015).

In England, the current Standards for Microbiology Investigations protocols are specific for the isolation of non-sorbitol fermenting colonies of *E. coli* serogroup O157 on cefixime tellurite sorbitol MacConkey (CT-SMAC) agar. STEC serogroups other than O157 (non-O157 STEC) are not detected using this method (Byrne et al., 2014). However, since 2012 the implementation of commercial PCR assays for the detection of STEC in faecal specimens from cases with symptoms of gastrointestinal infection, at a twelve local hospital laboratories, has resulted in an increase in the detection of non-O157 STEC (Byrne et al., 2014).

Faecal specimens that are PCR positive for the Shiga Toxin (*stx*) genes at the local hospital laboratories in England are sent to the Gastrointestinal Bacterial Reference Unit (GBRU) at Public Health England (PHE) for isolation of STEC (Jenkins et al., 2012) and subsequent serotyping (Gross and Rowe, 1985). Recent advances in whole genome sequencing (WGS) have led to the development of a method for high throughput sequencing of bacterial genomes at low cost (Joensen et al., 2014). During 2014, we evaluated the use of WGS for routine public health surveillance of non-O157 STEC by comparing this approach to phenotypic serotyping and PCR for subtyping the *stx*-encoding genes (Persson et al., 2007).

MATERIALS AND METHODS

All 167 strains of non-O157 STEC isolated during 2014 were phenotypically serotyped by the agglutination of antibodies raised in rabbits to the lipopolysaccharide O antigen and to the flagella H antigen (Gross and Rowe, 1985). Real-time PCR targeting *stx1* and *stx2* and the *stx* subtyping PCR was performed as previously described (Persson et al., 2007; Jenkins et al., 2012).

Genomic DNA extracted from 140 of the 167 strains of non-O157 STEC was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation Kits (Illumina) and sequenced using the Illumina HiSeq 2500. A reference database, *SerotypeFinder*, containing the gene sequences encoding the 180 O antigen groups (*wzx*, *wzy*, *wzm*, and *wzt*) and the 53 H antigens (*fliC*, *flkA*, *flaA*, *flmA*, and *fliA*) was constructed and developed by Joensen et al. (2015). Using the *GeneFinder* tool (Dounith unpublished), FASTQ reads were mapped to the genes in the *SerotypeFinder* database using Bowtie 2 (Langmead and Salzberg, 2012) and the best match to each of the O and H determinants was reported with metrics including coverage, depth, mixture and homology in an XML format for quality assessment. Only *in silico* predictions of serotype that matched to a gene determinant at >80% nucleotide identity over >80% length were accepted. *Stx* subtyping was performed as described by Ashton et al. (2015). FASTQ sequences were deposited in the National Center for Biotechnology Information Short Read Archive under the bioproject PRJNA248064.

RESULTS

Whole genome sequences were available for 140 of the 167 non-STEC isolates reported in 2014 (Supplementary Table). Of these, 102 had a phenotypically derived serogroup, 25 did not agglutinate with the antisera in the serotyping scheme raised to the known *E. coli* serogroups and were designated “O unidentifiable,” and 13 did not express the O antigen and were designated “O rough.” Of the 102 isolates where phenotypic and genotypic serotyping could be compared, 98 gave fully concordant results (Supplementary Table). The most common non-O157 STEC serogroups detected were O146 (22) and O26 (18). There were 15 strains of STEC O55, all from cases linked to an outbreak in the South of England.

Of the four results that were not fully concordant, two isolates serogrouped as O186 phenotypically but were designated O123/O186 by *in silico* serotyping and one typed as O178 phenotypically and was designated O153/178 using WGS data. There was one mismatch; STEC O74 was identified as STEC O187 when the serotype was derived from the genome. The *in silico* serotyping method failed to type one isolate, STEC O146:H21, due to the short read sequences having low mapping coverage of the O antigen encoding genes. All but one of the 38 isolates that could not be phenotypically serotyped (designated O unidentifiable or O rough) were serotyped using the WGS data (Supplementary Table). The most common WGS derived serotypes that were untypable using the phenotypic approach were O91:H14, O117:H7, and O80:H2.

There were 102 isolates that were processed for H-typing, of which 29 were found to be non-motile and could not be typed. Of the 73 isolates where a flagella type was available by traditional phenotypic typing, all results matched the H-type derived from the WGS data. All the non-motile isolates were typable using the *in silico* serotyped by *in silico* serotyping (Supplementary Table).

Of the 140 sequenced non-O157 isolates, 52 (37.1%) harboured *stx1* only, 42 (30.0%) had *stx2* only, 46 (32.9%) carried *stx1* and *stx2*. Of these, *stx* subtyping PCR results were available for 131 isolates and 121 of these had concordant results with the *stx* subtype derived from the WGS data (Supplementary Table).

The most frequently detected *stx* subtype profile was *stx1a* only (38), most commonly associated with serotypes O103:H2 (14), O26:H11 (8), and O117:H7 (8) (Supplementary Table). There were 14 isolates with *stx1c* only, including O146:H21 (5) and O76:H19 (3). All *stx1a* and *stx2b* strains belonged to STEC O91:H14 (9). There were 23 isolates that had *stx1c* and *stx2b* associated with serogroups O146:H21 (9), O128:H2 (8), O113:H4 (2), and O174:H8 (2). The combination of *stx1a* and *stx2a* was found almost exclusively in O26:H11 (7), the one exception being an isolate of STEC O71:H2. Of the 21 isolates harbouring *stx2a* only, 15 were STEC O55:H7 and two were O26:H11. Of the 10 isolates that had *stx2d*, three belonged to O146:H21 (and also had *stx1c*), six were O80:H2 and one was O26:H11. Other rare subtypes (one case of each) included *stx2e* (O8:H9) and *stx2g* (O187:H28) (Supplementary Table).

There were 10 cases of HUS in 2014 (including five cases belonging to an outbreak of STEC O55:H7), eight had STEC harbouring *stx2a*, one had STEC O80:H2 carrying *stx2d*, and O103:H2 *stx1a* was isolated from the tenth case.

DISCUSSION

The results of this study showed WGS provided a reliable and robust one-step process for characterisation of STEC. Previous studies have shown an increasing number of strains of STEC reported as “O group unidentifiable” due to antisera failing quality control procedures, unresolvable cross reactions, lack of expression of O antigens (designated “rough”) or novel serogroups (Jenkins et al., 2003; Byrne et al., 2014). In this study, all but one of the isolates that were previously phenotypically untypable, were serotyped using data derived from the genome.

Of the 10 mismatched results identified in the comparison between the stx subtyping PCR (Scheutz et al., 2012) and the WGS approach (Ashton et al., 2015), all 10 had additional stx subtypes detected by PCR that were not identified in the WGS data (Supplementary Table). Non-specific primer binding during PCR amplification, due to the similarity of the stx2 subtype gene sequences was the most likely cause.

Historically, the most common stx profile of STEC O26 was stx1a but over the last 10 years a more virulent STEC O26 variant harbouring *stx2a* has emerged across Europe (Bielaszewska et al., 2013). The enhanced strain characterisation data that WGS provides facilitated the surveillance of emerging strains of STEC associated with more severe disease (for example STEC O55:H7 stx2a, STEC O26:H11 stx2a, STEC O80:H2 stx2d) and with novel stx profiles (for example STEC O26:H11 stx2d) and enabled us to compare data with colleagues in the field (Mariani-Kurkdjian et al., 2014; Delannoy et al., 2015). Previous studies have also reported an association between stx2a and severe disease (Ethelberg et al., 2004; Byrne et al., 2014). The *eae* gene was detected in 62 (44%) of the 140 non-O157 STEC isolates. All 10 isolates from the HUS cases had *eae*. None of the STEC strains in this data set had *aggR*, previously detected in highly pathogenic STEC variants (Boisen et al., 2015).

Prior to the implementation of WGS, due to limited resources and time constraints, H-typing and stx subtyping of STEC were not routinely reported by GBRU. Deriving the full serotype from

WGS data in real time has enabled us to report a higher level of strain discrimination while stx subtyping provides data on the pathogenic potential of each isolate, enabling us to predict clinical outcome of each case and to monitor the emergence of hyper-virulent strains.

AUTHOR CONTRIBUTIONS

MW and NP isolated the STEC and performed the real-time stx PCR. AG and SL performed the phenotypic serotyping, stx subtyping PCR and extracted the DNA. CJ, MC, MW, and NP implemented the wet lab WGS pipelines and performed analysis. TD and PA implemented the bioinformatics pipelines and performed analysis. CJ, MC, and TD wrote the manuscript.

FUNDING

This work was supported by the National Institute for Health Research and Health Protection Research Unit in Gastrointestinal Infections at the University of Liverpool.

ACKNOWLEDGMENTS

We would like to acknowledge Yoshini Taylor and Dawn Hedges at GBRU, Rediat Twolde in the Infectious Disease Informatics Unit and Cath Arnold in the Genomics Service Unit at PHE. We would very much like to thank Flemming Scheutz at the Statens Serum Institute in Copenhagen and Katrine Joensen at the Danish Technical University for sharing the *SerotypeFinder* database with us.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00258>

REFERENCES

- Ashton, P. M., Perry, N., Ellis, R., Petrovska, L., Wain, J., Grant, K. A., et al (2015). Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. *PeerJ*. 3:e739. doi: 10.7717/peerj.739
- Bielaszewska, M., Mellmann, A., Bletz, S., Zhang, W., Köck, R., Kossow, A., et al (2013). Enterohemorrhagic *Escherichia coli* O26:H11/H-: a new virulent clone emerges in Europe. *Clin. Infect. Dis.* 56, 1373–1381. doi: 10.1093/cid/cit055
- Boisen, N., Melton-Celsa, A. R., Scheutz, F., O'Brien, A. D., and Nataro, J. P. (2015). Shiga toxin 2a and enteroaggregative *Escherichia coli* - a deadly combination. *Gut. Microbes*. 6, 272–278. doi: 10.1080/19490976.2015.1054591
- Byrne, L., Jenkins, C., Launders, N., Elson, R., and Adak, G. K. (2015). The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009–2012. *Epidemiol. Infect.* 143, 3475–3487. doi: 10.1017/S0950268815000746
- Byrne, L., Vanstone, G. L., Perry, N. T., Launders, N., Adak, G. K., Godbole, G., et al (2014). Epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England, 2009–2013. *J. Med. Microbiol.* 63, 1181–1188. doi: 10.1099/jmm.0.075895-0
- Delannoy, S., Mariani-Kurkdjian, P., Bonacorsi, S., Liguori, S., and Fach, P. (2015). Characteristics of emerging human-pathogenic *Escherichia coli* O26:H11 strains isolated in France between 2010 and 2013 and carrying the *stx2d* gene only. *J. Clin. Microbiol.* 53, 486–492. doi: 10.1128/JCM.02290-14
- Ethelberg, S., Olsen, K. E., Scheutz, F., Jensen, C., Schiellerup, P., Enberg, J., et al (2004). Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg. Infect. Dis.* 10, 842–847. doi: 10.3201/eid1005.030576
- Gross, R. J., and Rowe, B. (1985). “Serotyping of *Escherichia coli*,” in *The Virulence of Escherichia coli: Reviews and Methods (Special Publication of the Society for General Microbiology no. 13)*, ed M. Sussman (London: Academic Press), 345–360.
- Jenkins, C., Lawson, A. J., Cheasty, T., and Willshaw, G. A. (2012). Assessment of a real-time PCR for the detection and characterization of verocytotoxinogenic *Escherichia coli*. *J. Med. Microbiol.* 61, 1082–1085. doi: 10.1099/jmm.0.041517-0
- Jenkins, C., Willshaw, G. A., Evans, J., Cheasty, T., Chart, H., Shaw, D. J., et al (2003). Subtyping of virulence genes in verocytotoxin-producing *Escherichia coli* other than serogroup O157 associated with disease in the United Kingdom. *J. Med. Microbiol.* 52, 941–947. doi: 10.1099/jmm.0.05160-0

- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy *in silico* serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Mariani-Kurkdjian, P., Lemaître, C., Bidet, P., Perez, D., Boggini, L., Kwon, T., et al (2014). Haemolytic-uraemic syndrome with bacteraemia caused by a new hybrid *Escherichia coli* pathotype. *New Microbes New Infect.* 2, 127–131. doi: 10.1002/nmi2.49
- Persson, S., Olsen, K. E., Ethelberg, S., and Scheutz, F. (2007). Subtyping method for *Escherichia coli* shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J. Clin. Microbiol.* 45, 2020–2024. doi: 10.1128/JCM.02591-06
- Scheutz, F., Teel, L. D., Beutin, L., Piérard, D., Buvens, G., Karch, H., et al (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12

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

Copyright © 2016 Chattaway, Dallman, Gentle, Wright, Long, Ashton, Perry and Jenkins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Whole Genome Sequencing for Genomics-Guided Investigations of *Escherichia coli* O157:H7 Outbreaks

Brigida Rusconi^{1,2}, Fatemeh Sanjar^{1,2}, Sara S. K. Koenig^{1,2}, Mark K. Mammel³, Phillip I. Tarr⁴ and Mark Eppinger^{1,2*}

¹ South Texas Center for Emerging Infectious Diseases, University of Texas at San Antonio, San Antonio, TX, USA,

² Department of Biology, University of Texas at San Antonio, San Antonio, TX, USA, ³ Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, MD, USA, ⁴ Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Patrick Fach,
ANSES, France
Atsushi Iguchi,
University of Miyazaki, Japan

*Correspondence:

Mark Eppinger
mark.eppinger@utsa.edu

Specialty section:

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

Received: 06 March 2016

Accepted: 08 June 2016

Published: 30 June 2016

Citation:

Rusconi B, Sanjar F, Koenig SSK, Mammel MK, Tarr PI and Eppinger M (2016) Whole Genome Sequencing for Genomics-Guided Investigations of *Escherichia coli* O157:H7 Outbreaks. *Front. Microbiol.* 7:985. doi: 10.3389/fmicb.2016.00985

Multi isolate whole genome sequencing (WGS) and typing for outbreak investigations has become a reality in the post-genomics era. We applied this technology to strains from *Escherichia coli* O157:H7 outbreaks. These include isolates from seven North America outbreaks, as well as multiple isolates from the same patient and from different infected individuals in the same household. Customized high-resolution bioinformatics sequence typing strategies were developed to assess the core genome and mobilome plasticity. Sequence typing was performed using an in-house single nucleotide polymorphism (SNP) discovery and validation pipeline. Discriminatory power becomes of particular importance for the investigation of isolates from outbreaks in which macrogenomic techniques such as pulse-field gel electrophoresis or multiple locus variable number tandem repeat analysis do not differentiate closely related organisms. We also characterized differences in the phage inventory, allowing us to identify plasticity among outbreak strains that is not detectable at the core genome level. Our comprehensive analysis of the mobilome identified multiple plasmids that have not previously been associated with this lineage. Applied phylogenomics approaches provide strong molecular evidence for exceptionally little heterogeneity of strains within outbreaks and demonstrate the value of intra-cluster comparisons, rather than basing the analysis on archetypal reference strains. Next generation sequencing and whole genome typing strategies provide the technological foundation for genomic epidemiology outbreak investigation utilizing its significantly higher sample throughput, cost efficiency, and phylogenetic relatedness accuracy. These phylogenomics approaches have major public health relevance in translating information from the sequence-based survey to support timely and informed countermeasures. Polymorphisms identified in this work offer robust phylogenetic signals that index both short- and long-term evolution and can complement currently employed typing schemes for outbreak ex- and inclusion, diagnostics, surveillance, and forensic studies.

Keywords: *Escherichia coli*, O157:H7, EHEC, phylogenomics, outbreaks, single nucleotide polymorphism, genomic epidemiology, whole genome sequence typing

INTRODUCTION

Microbial pathogens with a foodborne etiology present major challenges to public health. *Escherichia coli* has been divided into different pathovars based on key virulence factors that define their pathogenicity (Sadiq et al., 2014). One particularly daunting pathovar among the Shiga toxin producing *E. coli* (STEC) are strains of the enterohemorrhagic O157:H7 serotype, which can be transmitted by a variety of vehicles, and causes serious human disease (Tarr et al., 2005). Currently, there is no effective treatment or prophylaxis for hemolytic uremic syndrome (HUS) (Goldwater and Bettelheim, 2012), and use of antibiotics is not indicated (Freedman et al., 2016). Since its discovery in 1982, this lineage has rapidly evolved from a rare serotype into the now globally dominant enterohemorrhagic *E. coli* (EHEC) serotype. A remarkable feature is its low infectious dose; it is estimated that 10–100 colony-forming units (CFUs) are sufficient to cause disease (Tilden et al., 1996; Tuttle et al., 1999). For the above reasons, prevention of human infection is critical, and early identification of outbreaks is highly worthwhile. However, only rudimentary information exists regarding the genomic heterogeneity that can be expected within outbreaks (STEC Outbreaks). Moreover, current typing schemes, such as pulse field gel electrophoresis (PFGE) and multiple locus variable number of tandem repeats analysis (MLVA), often lack the resolution to differentiate organisms that form tightly clonal phylogenetic clusters within the O157:H7 clade (Eppinger et al., 2011b; Turabelidze et al., 2013; Underwood et al., 2013; Rusconi and Eppinger, 2016). Additionally, PFGE is subject to technological and interpretation challenges (Davis et al., 2003).

Increasing technologic economies offer new opportunities for sequence-based typing of microbial pathogens for public health purposes (den Bakker et al., 2014; Joensen et al., 2014; Leekitcharoenphon et al., 2014; Holmes et al., 2015). While it would be ideal to refer a clinical strain's sequence to a reference, of the 445 publicly available genomes of *E. coli* O157:H7 and its close relative O55:H7 (O157:H7 Genomes) (Kulasekara et al., 2009; Zhou et al., 2010; Eppinger et al., 2011a, 2013; Sanjar et al., 2014, 2015), to date only 11 have been closed (Hayashi et al., 2001; Perna et al., 2001; Kulasekara et al., 2009; Zhou et al., 2010; Eppinger et al., 2011b, 2013; Kyle et al., 2012; Xiong et al., 2012; Latif et al., 2014; Sanjar et al., 2014, 2015; Cote et al., 2015). Whole genome sequencing (WGS) can provide the necessary resolution power to investigate apparent single source outbreaks (Eppinger et al., 2011b; Hasan et al., 2012; Turabelidze et al., 2013) because the granularity of WGS data provides considerable confidence in assigning like vs. not-like status to two potentially linked pathogens (Gilchrist et al., 2015). Such data can also link pathogens to vehicles or environmental isolates most precisely (Bentley and Parkhill, 2015). WGS can offer additional advantages: serotypes and virulence loci within pathogens can be identified (Scheut et al., 2012; Leekitcharoenphon et al., 2014; Lambert et al., 2015; Klemm and Dougan, 2016), and case management might theoretically be risk-optimized.

Optimization of *E. coli* O157:H7 sequence analysis methodologies depend on the scientific and epidemiologic inquiries and the data being analyzed. Pettengill et al. evaluated a number of single nucleotide polymorphism

(SNP) predicting tools and phylogenetic methodologies in prokaryotes and concluded that a reference-based approach, which accommodates missing data as well as infers phylogenetic reconstruction, is the most appropriate (Pettengill et al., 2014). Such a reference-based approach was recently used by the Alberta Provincial Laboratory for Public Health to study *E. coli* O157:H7 outbreaks together with virulence profiling and other molecular methods (Berenger et al., 2015). No specific virulence pattern distinguished the outbreak strains from sporadic strains (Berenger et al., 2015). Recent studies have expanded WGS typing to globally distributed strains and identified geographical genomic structuring based on distribution of *stx*-converting phage integration sites and SNPs (Mellor et al., 2015; Strachan et al., 2015) and provided a more detailed subtyping of *E. coli* O157:H7 (Griffing et al., 2015). However, clarity can also be gained by comparing closely related isolates to each other, rather than to reference strains (Leopold et al., 2009; Turabelidze et al., 2013).

Here we adapt WGS to a specifically developed SNP-based pipeline for the high resolution typing of *E. coli* O157:H7 by identifying SNPs within the core genome. In addition to SNP analysis in the core genome we assessed plasticity in the mobilome by LS-BSR and plasmid comparison (phages and plasmids) (Eppinger et al., 2011a,b, 2014; Hasan et al., 2012; Jenkins et al., 2015). We tested this pipeline on isolates from seven retrospectively analyzed EHEC O157:H7 outbreaks, six intra-household cases, and five clinical “plate-mate” pairs, i.e., colonies from the same primary isolation plate from the clinical laboratory.

MATERIALS AND METHODS

Strains in Study

We compared human isolates (Supplemental Table 1) of nine phylogenetic clades (Manning et al., 2008), so as to place the strains in the overall *E. coli* O157:H7 phylogenetic context. Strain-associated metadata of analyzed *E. coli* O157:H7 are provided in Supplemental Table 1. Outbreak strains were defined as a set of isolates from different cases of infection arising from a single point source, as determined by local health jurisdictions and/or the Centers for Disease Control and Prevention. Intra-household cluster strains were recovered from siblings within a household whose infections were not linked to a recognized outbreak. Because intra-household clusters could reflect co-primary infections rather than secondary transmission, we selected such pairings from among our strain set collection (Cornick et al., 2002; Besser et al., 2007) on the basis of prolonged intervals (4–6 days) between cases, so as to increase the likelihood that genomic diversity might emerge secondary to inter-host transmission. Plate mates are pairs of isolates from the same sorbitol-MacConkey agar plate used in clinical laboratories to diagnose the infection.

Bioinformatic Analyses for Polymorphisms Discovery in Core Genome and Mobilome

Developed bioinformatics workflows, methods and principles for SNP discovery and core and accessory genome analyses

performed in this study are described in **Figure 1** with external tools referenced in the legend. Multinucleotide insertions and deletions of polymorphic bases were not considered SNPs. To classify SNPs we mapped the annotation from the *de novo* annotated references with PROKKA and Prodigal ORF prediction (Hyatt et al., 2010), or the deposited annotation for EC4115 (Eppinger et al., 2011b). The core genome was defined as the set of genic and intragenic regions that were not repeated, did not contain phages, IS elements, or plasmid regions. Briefly for SNP discovery, reads were aligned with Bowtie2 (Langmead and Salzberg, 2012) to designated reference genomes. Resulting alignments were processed with FreeBayes (Garrison and Marth, 2012) with the following threshold settings: mapping quality 30, base quality 20, coverage 30, and allelic frequency 0.9. To account for false positive calls we used several SNP curation strategies: (i) Reference reads were mapped against the reference genome and false positives were identified by FreeBayes with the settings described above; (ii) If reads were not available, the post-assembly workflow created a reference-based NUCmer alignment and extracted SNPs with delta-filter and show-snps distributed with the MUMmer package (Delcher et al., 2003). SNP occurring in the excluded regions were removed. Cataloged SNPs from each genome were merged into a single SNP panel, and allelic status and chromosomal position were recorded. Curated SNPs were further processed by extracting the surrounding nucleotides (40 nt) and blastn against the query genomes (Altschul et al., 1990). Resulting alignments were parsed to remove SNP locations derived from ambiguous hits (≥ 2), non-uniformly distributed regions, and insertion or deletion events, as previously described (Myers et al., 2009; Morelli et al., 2010; Eppinger et al., 2011b, 2014; Vogler et al., 2011; Hasan et al., 2012).

Optical Maps

Optical mapping facilitated accurate phage profiling (Kotewicz et al., 2008). In total 12 maps were generated (Supplemental Table 1), either prepared by OpGen or contributed by FDA (Eppinger et al., 2011b, 2014). After gentle lysis and dilution, the extracted genomic DNA molecules from each strain were spread and immobilized onto derivatized glass slides. The genomic DNA was then digested with BamHI restriction enzyme maintaining the DNA fragment order. Using the ArgusTM Instrument, the DNA fragments were stained with YOYO-1 fluorescent dye and photographed using a fluorescent microscope interfaced with a digital camera. The optical data was converted to digital data, which defines single molecule restriction maps. Physical maps were complemented with *in silico* maps of other outbreak strains, and comparatively analyzed in MapSolverTM Optical Map Analysis software (Latreille et al., 2007; Zhou et al., 2007).

SNP PCR Validation

SNPs in four isolates from two outbreaks for which we possessed archived cultures were subjected to PCR confirmation using primer pair (89750-F 5'-ACA ACG ATA TGA TCG ACC AGC, 89750-R 5'-TTG TAC AGA AGA CCA TGC TCG) and (27005-F 5'-AGA GTA CGG ATT CAC CTT GCC, 27005-R 5'-AGT CAG GCA ATT CCT CGT GG, 78298-F 5'-AGT CAT TAC CAG GAA CAG CAG 78298-R 5'-TGT TCG AGA TTC TGG TGA

GTG) for strains from the Battle Ground Lake and Finley School District outbreak, respectively. Resulting amplicons were Sanger sequenced.

Multi Drug Resistance (MDR) Profiling

Susceptibility to amikacin, ampicillin, amoxicillin-clavulanic acid, cefoxitin, ceftiofur, ceftriaxone, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfisoxazole, tetracycline, and trimethoprim-sulfamethoxazole was assessed at FDA according to the NARMS methodology and manufacturer's instructions with the Sensititre automated system (Trek Diagnostic Systems, Westlake, OH) (Zhao et al., 2008). Resistance was determined by comparing MICs to Clinical and Laboratory Standards Institute (CLSI) values (Institute, 2013).

RESULTS AND DISCUSSION

Epidemiology of Investigated Strains

We analyzed 36 strains from seven US outbreaks as recognized by the CDC that occurred between 1998 and 2009 (Supplemental Table 1): (1) 11 children were infected after consumption of contaminated ground beef tacos in the Finley School District (FS) in 1998; (2) 28 swimmers at Battle Ground (BL) Lake State Park, WA, and eight secondary cases were infected in 1999; (3) 81 cases were attributed to lettuce served at multiple outlets of a taco chain (Taco John) in 2006; (4) 71 people were infected in a multistate outbreak after eating at Taco Bell (TB) in 2006; though the vehicle was not identified (Taco Bell); (5) 21 infections were attributed to a prolonged multi-state outbreak linked to the consumption of Totino's or Jeno's contaminated pepperoni pizza (Totino's pizza) (TP) in 2007; (6) 76 cases were attributed to a nationwide outbreak of contaminated cookie dough (CD) (Cookie dough) in 2009; (7) 26 patients from eight states were infected by beef traced to Fairbank Farms (FF) in 2009 (Fairbank Farms).

We further studied (#12) strains from six intrahousehold illnesses (IH), in which the pathogen probably spread between patients based on the long intervals between onset in the individual family members (Supplemental Table 1). Though we cannot exclude the possibility of infection from the same source (co-primary). The median incubation period of *E. coli* O157:H7 infections is 3 days (Bell et al., 1994), and onsets ranged between 4 and 6 days. We also studied pairs of isolates from the same primary plate in the clinical laboratory (plate-mates, PM) from six patients (**Figure 2**). The clinical strains were compared to strains representing the nine phylogenetic clades reported by Manning et al. (2008) (**Figure 2**, Supplemental Table 1).

Core Genome Phylogeny

We applied WGS typing strategies to determine the phylogenetic relatedness of the individual outbreak strains in the context of outbreak etiology, and to place them into the larger phylogenomic framework of the *E. coli* O157:H7 lineage (Leopold et al., 2009; Eppinger et al., 2011b; Dallman et al., 2015; Holmes et al., 2015; Jenkins et al., 2015). Of the 3313 SNPs identified in these 70 genomes, 2797 were intragenic and 516

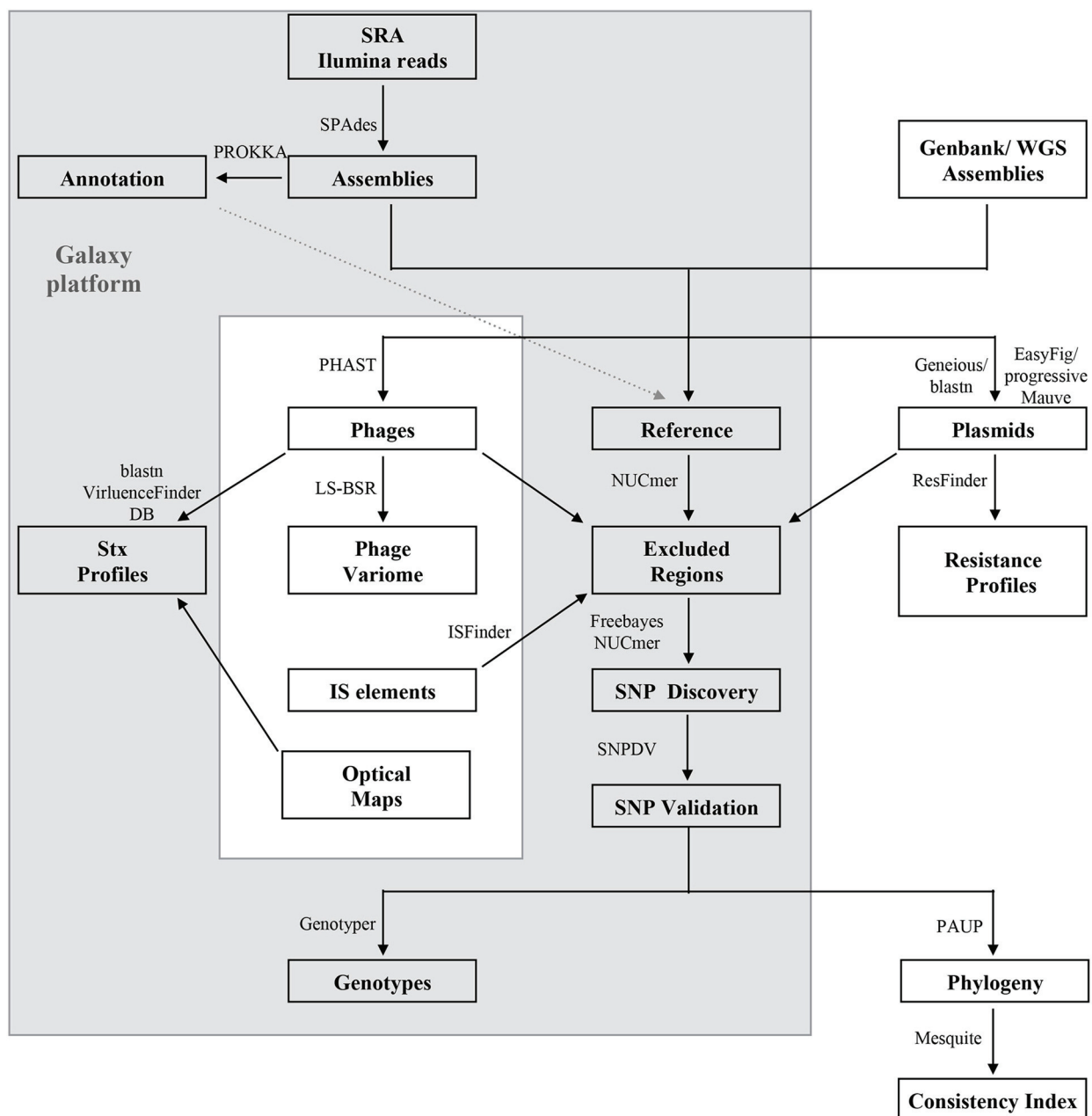


FIGURE 1 | Bioinformatics analyses of the *E. coli* O157:H7 core and mobile genome. The majority of the analyses were performed on Galaxy (Goecks et al., 2010), an open-source web-based bioinformatics platform (gray box). Illumina reads, draft or closed genomes were retrieved from the NCBI SRA or WGS sequence repositories (Supplemental Table 1). Reads were assembled with SPAdes v 3.5.0 (Bankevich et al., 2012), and annotated with PROKKA v 1.11.0 (Seemann, 2014). A reference genome was selected for each outbreak analysis and distinct regions were excluded from SNP validation as follows: Repeats with NUCmer (Delcher et al., 2003), phages with PHAST (Zhou et al., 2011) plasmids by querying against a custom *E. coli* plasmid database and IS elements from ISFinder (Siguiet et al., 2006). Optical maps were generated to facilitate accurate prophage and mobilome profiling (Eppinger et al., 2011b). Read based SNP discovery was based on Bowtie2 alignment (Langmead and Salzberg, 2012) and subsequent Freebayes SNP calling (Garrison and Marth, 2012) and for closed or draft genomes on NUCmer (Delcher et al., 2003), followed by blast- and PCR-based SNPs validations to curate for false-positives with strategies detailed in (Eppinger et al., 2011b, 2014). The identified SNPs were genotyped and used for phylogenetic reconstruction by maximum parsimony of the outbreaks with PAUP v4.0a146 (Wilgenbusch and Swofford, 2003). Consistency index and majority consensus trees were built with Mesquite (Maddison and Maddison, 2015). The mobilome was analyzed in regards to phage content with large scale-blast score ratio (LS-BSR) (Sahl et al., 2014) and computed matrices visualized with MeV (Saeed et al., 2003). Shiga toxins were identified by discontinuous megablast (Buhler, 2001; Ma et al., 2002) against the VirulenceFinder (Joensen et al., 2014) database. Plasmids were identified after alignment of draft genomes to the closest closed reference plasmid in Geneious (vR9) (Kearse et al., 2012). Unmapped remaining contigs were queried against NCBI nr plasmids by discontinuous megablast (Buhler, 2001; Ma et al., 2002). Identified plasmid homologs were compared and visualized with EasyFig (Sullivan et al., 2011) by tblastx (Altschul et al., 1990) or progressiveMauve (Darling et al., 2010). Plasmids and chromosomes were further surveyed for presence of resistance loci using ResFinder (Kleinheinz et al., 2014).

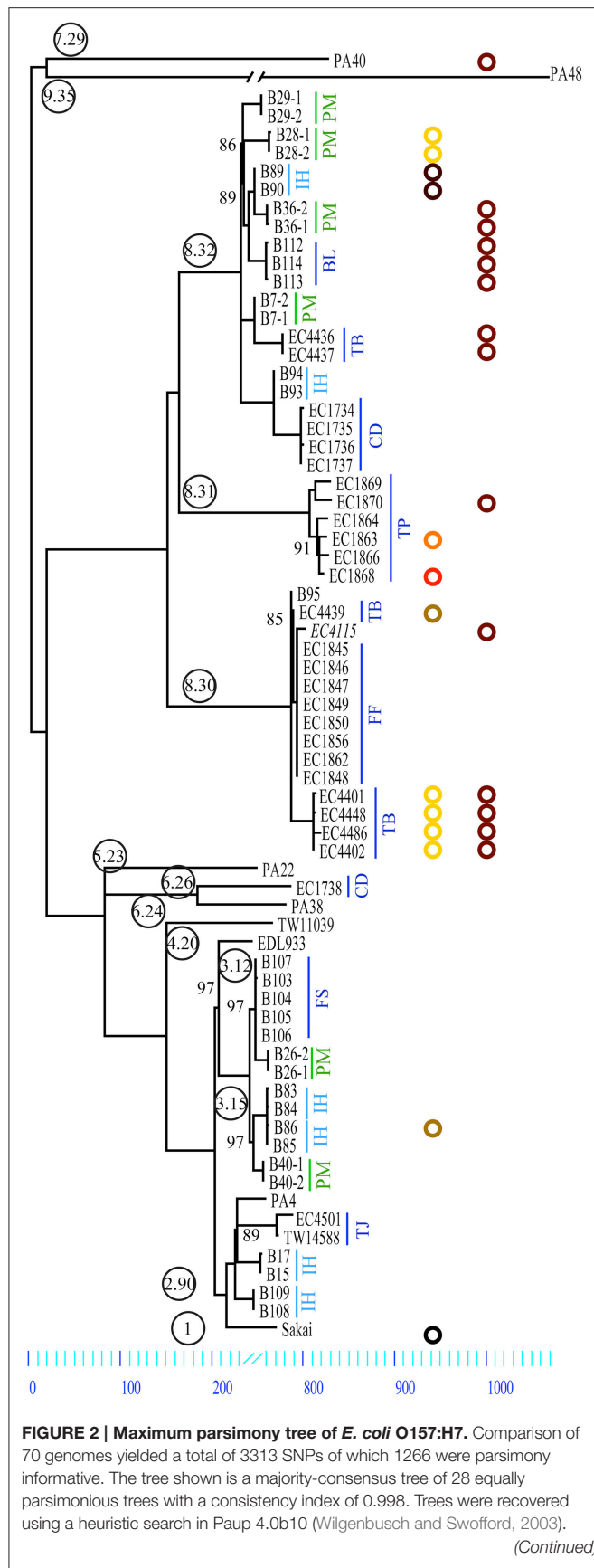


FIGURE 2 | Continued

Only nodes with bootstrap values below 100 are listed. Phylogenetic clade association is provided in circled numbers (Manning et al., 2006). Strains investigated are comprised of PM, Plate mates; IH, intrahousehold infections; BL, Battle Ground Lake; CD, Cookie Dough; FS, Finley School District; FF, Fairbank Farms; TB, Taco Bell; TJ, Taco John; and TP, Totino's Pizza. Our plasmid survey confirmed that all strains carry the lineage-specific plasmid pO157. Further we identified other plasmids with a size range from 34 to 78 kb. Plasmid type prevalence is represented in colored circles: p78 (yellow), p63 (orange), p55 (light brown), p39 (red), p36 (brown), p34 (dark brown), and a small 3.3 kb plasmid (black) (Makino et al., 1998). Plasmid p36 is homologous to pEC4115 (Eppinger et al., 2011b).

were intergenic (Table 1, Supplemental Dataset 1). We observed significantly more SNPs in intergenic regions (chi square test, $p < 10^{-14}$) than would be expected when considering the average intergenic frequency in EC4115 of 11.1% when compared to the percentage (15.5%) delineated from the cataloged SNPs. We note that, even though we excluded repeated regions and phages/mobilome during SNP discovery, thereby reducing the genome content by 20%, the coding to non-coding ratio of the remaining core genome remained stable. Homoplasy was negligible: only seven homoplastic SNPs were found dispersed throughout the chromosome (Supplemental Table 2) evidenced by a consistency index of 0.998. Of the seven homoplastic SNPs four are in *rpoS*, which is known to be highly polymorphic in *E. coli* O157:H7 (Uhlich et al., 2013). In line with our previous findings (Leopold et al., 2009; Eppinger et al., 2011a,b), SNPs were evenly distributed throughout the chromosome (Figure 2) without any mutational hot spots as found in other enteric pathogens (Hasan et al., 2012; Eppinger et al., 2014). From the cataloged SNP panel we delineated a total of 77 individual SNP genotypes. These genotypes represent only two-thirds of the 115 nodes in the tree (Figure 2, Supplemental Table 2), which can be attributed to the lack of terminal strain-specific SNPs (Figure 2). Among the cataloged 3313 SNPs, approximately one-third (#1266) is parsimony-informative. The SNPs in PA40 and PA48 are not strain-specific, but indicate the relative phylogenetic distance that separates these clade 7 and 9 strains from the other clades (Manning et al., 2008). As evidenced in the tree topology, approximately half of the parsimony non-informative SNPs (#1046) is introduced by reference strain PA48 from clade 9 (Figure 2, Supplemental Table 2). Among investigated strains PA48 is phylogenetically closest to the progenitor O55:H7 serotype (Feng et al., 1998; Manning et al., 2008; Zhou et al., 2010). This clade is also within the most ancient cluster of *E. coli* O157:H7 (Leopold et al., 2009) and higher SNP counts are indicative of more time to accrue mutations than in other phylogenetic groups that have emerged more recently.

Genomic Epidemiology of North American Outbreaks

Guided by the established phylogenomic framework (Figure 2), we analyzed outbreak specific “genome” characteristics and polymorphic heterogeneity in seven different North American outbreaks using a common (EC4115), as well as outbreak-specific references. We specifically applied this dual

TABLE 1 | SNPs characteristics for 70 *E. coli* O157:H7.

	Total	NSYN	NSYN/ NSYN	SYN	SYN/SYN	Intergenic	Genic
SNPs	3313	1712	1	1083	1	516	2797
NI	2048	1054	1	656	1	335	1712
PI	1266	658	0	427	0	181	1085
Genes	1821	1264	1	891	1	0	2157
Stop gain	50	57	0	0	0	0	57
Stop loss	16	17	0	0	0	0	17
Hypothetical proteins	245	269	1	103	0	0	373
Transition	2254	1037	1	852	1	363	1891
Transversion	1064	676	1	231	1	155	909
Multiallelic	5	1	1	2	1	2	5

reference genome approach to improve resolution power by enabling polymorphism discovery in parts of the core genome integral to the outbreak-associated strains, but not necessarily present in a more phylogenetic distant reference like EC4115 (Figure 2). We note here that our investigation of the 2006 Spinach (SP) outbreak revealed a number of subtle polymorphisms distinguishing all the recovered Maine isolates from the remainder of SP strains. Such subtle polymorphisms would have clearly evaded detection by using a reference from outside the SP outbreak (Eppinger et al., 2011b). In general we observed limited plasticity among related outbreak strains when compared to the closed reference genome EC4115 (Tables 2–5) (Eppinger et al., 2011b). Strains with increased SNP numbers were either from cases that were epidemiologically predicted to be outliers, or that could not be read-corrected. For example, the TB outbreak associated strains included four strains classified by CDC as temporal outliers (Supplemental Table 1), two (EC4436, EC4437) separated by 285 SNPs and one (EC4439) by 27 SNPs from the core outbreak cluster (Figure 2, Tables 2, 3). According to our SNP analysis, the remaining strain EC4448 should be considered to be derived from a single point source, even if this isolate is separated by a single homoplasic nonsynonymous (nsyn) SNP (Figures 2, 3, Tables 2, 3, Supplemental Table 2). Notably, *rpoS* carries this homoplasic stop codon mutation, which is known to be highly polymorphic in *E. coli* O157:H7; particularly in regards to premature stop codons that affect curli expression and biofilm formation (Uhlich et al., 2013). The same SNP was identified with an outbreak specific reference EC4401 in addition to multiple (#31) reference specific alleles (Figure 4A, Tables 2, 3, Supplemental Dataset 1). These SNPs were mainly located in intergenic (#30) regions and probably caused by over-predictions because of a lack of reads in the genome repository, and consequently inability to perform quality control. We observed the same phenomenon of over-prediction for the TJ outbreak strains, separated by 24 SNPs (Tables 2, 3, Supplemental Dataset 1); again no read data were available to us. We found the majority of predicted SNPs clustered mainly in close proximity either in intergenic regions or within the boundaries of the same gene, indicative of low quality sequence regions. Intragenic SNPs were identical to those found in EC4115,

except for two additional SNPs in the *lac* repressor and in *rpoS* (Supplemental Dataset 1). The CD outbreak set underwent both contig and read-based discovery, which again over-predicted SNPs for EC1734 (no reads) due to a lack of reads for quality control (Figures 2, 4B, Tables 2, 3). Moreover, the production lot isolate EC1738 was placed on a distant branch (clade 6.26), separated from all human isolates tightly clustered in clade 8.30 (Figure 1, Table 2, Supplemental Table 1). Hence, we consider this strain as an outlier, which is phylogenetically unrelated to the case isolates. Among the outbreak-specific SNPs we detected one synonymous (syn) and two nsyn SNPs in EC1736, but the syn was also detected using EC4115 as a reference (Table 2, Supplemental Dataset 1). Archived strains were not available for this outbreak and we could therefore not confirm if EC1736 truly carries these 3 SNPs, which would question its inclusion into the outbreak.

For three outbreaks (FF, FS, and BL) we identified only a single or no SNP when referenced to EC4115 (Figure 2, Tables 2, 3). A single intergenic and three nsyn mutations were identified when using an FF outbreak-specific reference strain EC1856 (Tables 2, 3). The three nsyn SNPs did not affect domain prediction in Pfam (Finn et al., 2016). B112 of the BL outbreak had a syn SNP in a tRNA-histidine ligase (#3460738) not found in any other *E. coli* O157:H7 genome deposited (nr or WGS). This SNP was identified in both instances when using EC4115 or an outbreak-specific reference (Supplemental Dataset 1). This SNP was confirmed using PCR amplicon sequencing. Using the alternative FS outbreak-specific reference one intergenic SNP in B105 and one nsyn SNP affecting the Nitrogen regulation protein NR(I) (ECH74115_RS26390) (B107 and B105) were identified (Tables 2, 3). SNP discovery predicted an outbreak-specific allele in three strains. However, these SNPs are false positives, as they could not be confirmed by PCR sequencing. The SNP (#693920) in FS strain B103 was identified as false-positive homoplasic SNP also observed in plate mate and intrahousehold strains with an allelic frequency below 0.9 (Supplemental Table 2). During SNP prediction we identified 52 SNPs in strain B103 that were not found in the other FS outbreak strains. These 52 SNPs were all located in a phage region that corresponds to the tandem integrated SP1/2 phages (Hayashi et al., 2001). The SNPs were all false-positives due to the presence of an additional phage in B103 related to a prophage from organism pro483 (NC_028943) (Supplemental Figure 3). The tail fiber proteins of these two phages were sufficiently similar to misalign reads for B103. This exemplifies the importance of SNP curation and assessment according to the genomic region in which they originate, as independent horizontal acquisition of segments can introduce epidemiologically misleading SNPs (Pettengill et al., 2014), also known as epidemiological type 2 errors of attribution.

The TP outbreak strains revealed a highly distinct SNP pattern compared to the genomic plasticity reported for other outbreaks (Figures 1, 5, Tables 2, 3). Two distinct phylogenetic clusters separated by 16 SNPs were observed. Additionally, each strain carried at least 4–17 strain-specific SNPs. Comparison to outbreak-specific reference EC1863 confirmed the relative high number of strain-specific SNPs (Tables 2, 3, Figure 5). In

TABLE 2 | Comparison of common vs. outbreak-specific reference genic SNPs.

Isolate	Sequences analyzed ^a	Synonymous SNPs in isolates, among backbone ORFs			Nonsynonymous SNPs in isolates, among backbone ORFs	Comments
		Compared to reference EC4115	Compared to outbreak strain	Compared to reference EC4115		
FINLEY SCHOOL OUTBREAK, ALL ISOLATES CLADE 3.15						
B103	A	0	0	0	1	All co-primary cases
B104	A	0	0	0	1	
B105	A	0	Reference	0	Reference	
B106	A	0	0	0	0	
B107	A	0	0	0	1	
TACO BELL OUTBREAK, ALL CLADE 8.30 EXCEPT EC4436-7 CLADE 8.32						
EC4401	B	1	Reference	0	Reference	Eppinger et al., 2011b
EC4402	A	0	0	0	0	Case isolate
EC4436	A	90	Excluded	155	Excluded	Temporal outliers
EC4437	A	90	Excluded	155	Excluded	
EC4439	A	5	Excluded	17	Excluded	
EC4448	A	0	0	1	2	
EC4486	B	0	1	4	4	
TACO JOHN, ALL ISOLATES CLADE 2.90						
EC4501	B	1	0	6	8	Eppinger et al., 2011b
TW14588	C	0	Reference	0	Reference	
FAIRBANK FARMS OUTBREAK, ALL ISOLATES CLADE 8.30						
EC1845	A	0	0	0	2	Case isolate
EC1846	A	0	0	0	2	
EC1847	A	0	0	0	2	
EC1848	A	0	0	0	2	
BATTLEGROUND LAKE OUTBREAK, ALL ISOLATES CLADE 8.32						
B112	A	1	1	0	0	Case isolates
B113	A	0	0	0	0	
B114	A	0	Reference	0	Reference	
COOKIE DOUGH OUTBREAK, EC1738 CLADE 6.26, EC1734-7 CLADE 8.32						
EC1738	B	168	excluded	298	excluded	Product isolate
EC1734	B	0	5	2	5	Case isolates
EC1735	A	0	1	0	2	
EC1736	A	1	Reference	1	Reference	
EC1737	A	0	1	0	2	
TOTINO'S, PIZZA OUTBREAK ALL ISOLATES CLADE 8.31						
EC1863	A	6	Reference	11	Reference	Case isolates
EC1864	A	9	10	10	8	
EC1866	A	6	7	12	8	
EC1868	A	6	3	6	6	
EC1869	A	7	15	9	21	
EC1870	A	10	20	9	21	
FAIRBANK FARMS OUTBREAK, ALL ISOLATES CLADE 8.30						
EC1849	A	0	0	0	2	Case isolate
EC1850	A	0	0	0	2	
EC1856	A	0	Reference	0	Reference	
EC1862	A	0	0	0	3	

^a This study/short reads in NCBI = A, WGS = B, assembled genomes in NCBI = C.

contrary to our observations for strain-specific SNPs in the above discussed outbreaks, these SNPs are neither concentrated in specific regions nor more frequent in intergenic than in genic

regions (Tables 2, 3). The EC1869/EC1870 branch contributes roughly 60% of all SNPs (Supplemental Dataset 1). Based on the established phylogenetic topology we hypothesize that two

TABLE 3 | Comparison of common vs. outbreak-specific reference intergenic SNPs.

	Isolate	Sequences analyzed ^a	SNPs in isolates, in intergenic regions		Comments ^b
			Compared to reference EC4115	Compared to outbreak strain	
FS	B103	A	1	1	Reference specific allele
	B104	A	0	1	
	B105	A	0	Reference	
	B106	A	0	1	
	B107	A	0	1	
BG	B112	A	0	0	
	B113	A	0	0	
	B114	A	0	Reference	
TB	EC4401	B	1	Reference	F
	EC4402	A	0	27	
	EC4436	A	40	Excluded	
	EC4437	A	40	Excluded	
	EC4439	A	5	Excluded	
	EC4448	A	0	27	
	EC4486	B	3	31	
TJ	EC4501	B	8	16	F
	TW14588	C	1	Reference	
CD	EC1738	B	67	Excluded	F
	EC1734	B	0	4	
	EC1735	A	0	0	
	EC1736	A	0	Reference	
	EC1737	A	0	0	
TP	EC1863	A	2	Reference	High diversity
	EC1864	A	1	5	
	EC1866	A	2	6	
	EC1868	A	2	5	
	EC1869	A	10	14	
	EC1870	A	6	11	
FF	EC1845	A	0	1	Reference specific allele
	EC1846	A	0	1	
	EC1847	A	0	1	
	EC1848	A	0	1	
	EC1849	A	0	1	
	EC1850	A	0	1	
	EC1856	A	0	Reference	
	EC1862	A	0	1	

^a This study/short reads in NCBI = A, WGS = B, assembled genomes in NCBI = C.
^b F = mixed analysis with reads missing for some strains.

closely related but different *E. coli* O157:H7 contaminated a common vehicle, if, indeed, all cases had the same exposure. Two-thirds of the SNPs were strain-specific, denoting a particular high diversity within this outbreak (Figures 1, 5, Tables 2, 3).

Such a degree of genomic plasticity among epidemiologically linked strains has rarely been observed in *E. coli* O157:H7. Several scenarios could have led to this radial expansion: (i) the epidemiology linked cases together that actually were from different simultaneous outbreaks, (ii) the SNPs identified *in silico* are false positives and only PCR-confirmation could really confirm the true distance among the strains, (iii) the high rate of accumulated SNPs could be caused by a mutator genotype resulting in the accumulation of mutations in a short time span, (iv) the heterogeneity could be related to the protracted duration of the outbreak (3 months), vs. single, brief, single source-exposures as in the FS outbreak, or (v) heterogeneity caused by increased strain mutation rates during outbreaks as have been discussed for other enterics (Morelli et al., 2010). In support of our findings, Dallman et al. noted correlations between the length of the strain collection intervals and respective numbers of SNPs observed (Dallman et al., 2015).

The clonal nature of *E. coli* O157:H7 outbreaks was confirmed in the majority of the outbreak strains analyzed here, consistent with prior findings from SNP typing in other O157:H7 outbreaks (Turabelidze et al., 2013; Dallman et al., 2015; Holmes et al., 2015; Jenkins et al., 2015; Munns et al., 2016). We found the number of SNPs to be inversely proportional to the availability of reads. This highlights the critical importance of quality control for accurate SNP discovery by accounting for both underlying sequence quality and evolutionary context of the SNP carrying loci to curate for false-positives. In this regard, the relevance of excluding mobile regions when inferring outbreak relatedness is evidenced in the loss of at least two thirds of predicted SNPs that if considered would impair phylogenetic accuracy.

WGS Typing of Plate Mates Recovered from Human Infections

In the medical praxis typically a single colony is retrieved from a primary isolation plate and sent for further molecular analysis. It is therefore not clear how much genotypic diversity exists among infecting isolates of *E. coli* O157:H7 as shed from the same individual in a single stool. To answer this question, plate-mates (pairs of colonies) were separately saved from five patients (Figure 2, Supplemental Table 1) enrolled in a multi-state study of *E. coli* O157:H7 infections (Wong et al., 2012). In the EC4115 reference-based discovery, two PM possessed the same homoplastic intergenic SNP (Figure 2, Tables 4, 5), which was not confirmed after allelic verification. When using an internal reference these strains were undistinguishable. The results are in accordance with those of Dallman et al., who reported 0–2 SNPs among same patient isolates, with most (70%) having no SNP differences at all (Dallman et al., 2015). Our results from this limited study, therefore, point toward infection with a single *E. coli* O157:H7 clone as the underlying cause for the majority of infections. We previously reported that a single laboratory passage can produce SNPs in *E. coli* O157:H7, but SNPs arise only rarely (Eppinger et al., 2011b). In the course of naturally acquired human infections, our

TABLE 4 | Comparison of common vs. PM/IH-specific genic SNPs.

Isolate	Sequences analyzed ^a	Synonymous SNPs in isolates, among backbone ORFs		Nonsynonymous SNPs in isolates, among backbone ORFs	
		Compared to reference EC4115	Compared to outbreak strain	Compared to reference EC4115	Compared to outbreak strain
PM	B26-1	A	0	0	0
	B26-2	A	0	Reference	Reference
PM	B28-1	A	0	Reference	Reference
	B28-2	A	0	0	0
PM	B29-1	A	0	0	0
	B29-2	A	0	Reference	Reference
PM	B36-1	A	0	0	0
	B36-2	A	0	Reference	Reference
PM	B40-1	A	0	Reference	Reference
	B40-2	A	0	0	0
PM	B7-1	A	0	0	0
	B7-2	A	0	Reference	Reference
IH	B83	A	0	0	0
	B84	A	0	Reference	Reference
IH	B85	A	0	Reference	Reference
	B86	A	0	0	0
IH	B89	A	0	0	0
	B90	A	0	Reference	Reference
IH	B93	A	0	Reference	Reference
	B94	A	0	0	0
IH	B15	A	0	0	0
	B17	A	0	Reference	Reference
IH	B108	A	0	0	0
	B109	A	0	Reference	Reference

^aThis study/short reads in NCBI = A, WGS = B, assembled genomes in NCBI = C.

data endorse that *E. coli* O157:H7 SNPs are exceptionally rare events.

WGS Typing of Strains from Intrahousehold Infections

To determine if genomic changes in infecting *E. coli* O157:H7 occur during probable intrahousehold (IH) transmission, we analyzed a cohort of six pair isolates from IH infections where onset was quite delayed between cases (Figure 2). As with the PM pairs, EC4115 based SNP discovery resulted only in false positive homoplastic intergenic SNPs (Figure 2, Tables 3, 4) that were absent in the pair-wise analysis. Dallman et al. observed similar SNP distributions in household transmission cases in the UK, with 40% having no such differences in the core genome (Dallman et al., 2015). Interestingly, two IH cases of clade type 3.15 clustered together (Figure 2). A single syn SNP was

specific to the B83/B84 cluster. These cases were all from the same state and occurred in the same year, but epidemiological investigations suggest they are separate cases of IH transmissions with over 6 weeks between occurrence and 80 miles distance between the zip codes in which the cases resided. This application of WGS typing analysis can genomically link clusters that were not previously identified epidemiologically (Dallman et al., 2015).

In general the frequency of SNPs in intergenic and genic regions were similar, highlighting the random nature of SNPs identified. While there is clearly no applicable universal gold standard or criteria for outbreak ex- or inclusion in regards to SNP matrix distances, we note that a number of outbreak investigations have found between four to seven SNPs among strains with putative epidemiological links (Underwood et al., 2013; Joensen et al., 2014; Dallman et al., 2015; Holmes et al.,

TABLE 5 | Comparison of common vs. PM/IH-specific intergenic SNPs.

	Isolate	Sequences analyzed ^a	SNPs in isolates, in intergenic regions		Comments ^b
			Compared to reference EC4115	Compared to outbreak strain	
PM	B26-1	A	0	0	
	B26-2	A	0	Reference	
PM	B28-1	A	1	Reference	Homoplastic FP
	B28-2	A	0	0	
PM	B29-1	A	0	0	
	B29-2	A	0	Reference	
PM	B36-1	A	1	0	Homoplastic FP
	B36-2	A	0	Reference	
PM	B40-1	A	0	Reference	
	B40-2	A	0	0	
PM	B7-1	A	0	0	
	B7-2	A	0	Reference	
IH	B83	A	1	0	Homoplastic FP
	B84	A	0	Reference	
IH	B85	A	0	Reference	Homoplastic FP
	B86	A	1	0	
IH	B89	A	0	0	
	B90	A	0	Reference	
IH	B93	A	0	Reference	
	B94	A	0	0	
IH	B15	A	0	0	Homoplastic FP
	B17	A	1	Reference	
IH	B108	A	0	0	
	B109	A	0	Reference	

^aThis study/short reads in NCBI = A, WGS = B, assembled genomes in NCBI = C.
^bFalse positive = FP.

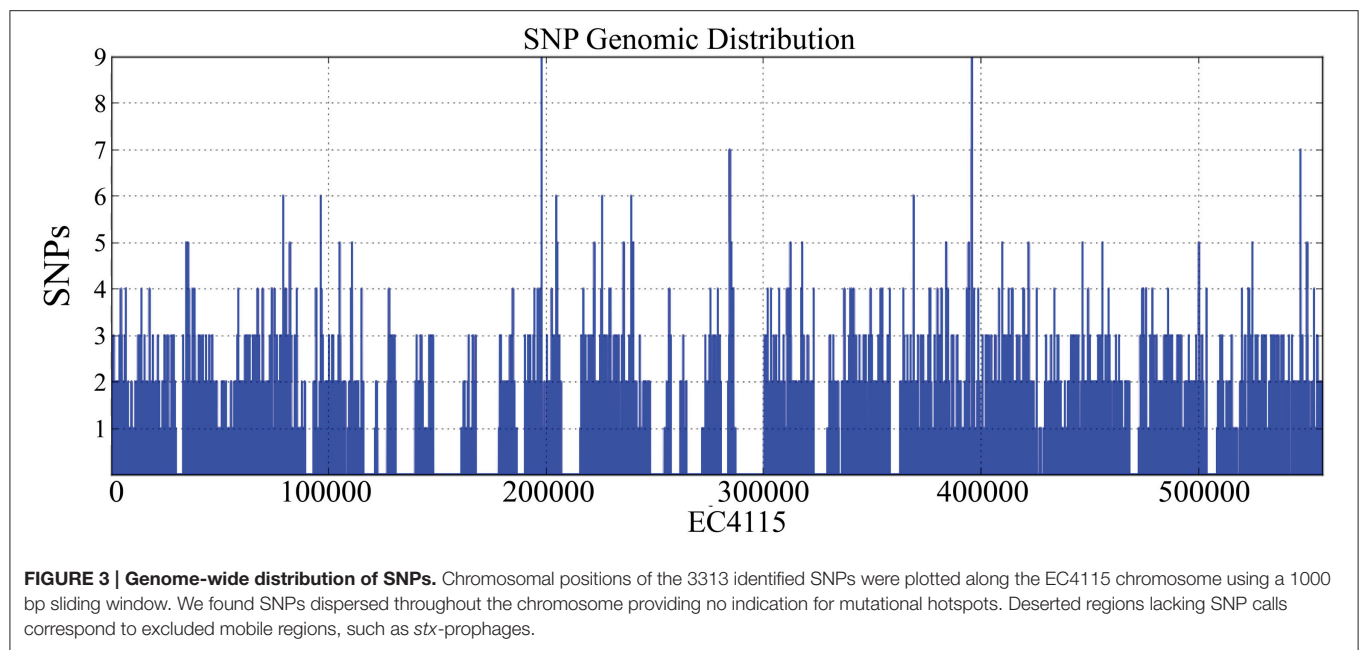
2015). However, these analyses are limited by only including the genic portions of the genomes and/or did not use an outbreak-specific reference for SNP discovery. This prevents identification of variations in parts of the core genome that are unique to outbreak-associated strains and not necessarily present in a distantly related closed reference strain. Moreover, only few studies use confirmatory PCR or other resequencing to validate *in silico* delineated SNPs (Eppinger et al., 2011b; Underwood et al., 2013).

Phage Profiles of Clinical U.S. Strains

The abundance of lambdoid phages in the EHEC O157:H7 genome hinders assembly of phage regions based on short reads alone (Eppinger et al., 2011b). Contig breaks often occur within the phage borders due to the conserved nature of structural

and replication proteins and hinder individual phage-level comparisons in the fragmented phage assemblies. Therefore, we applied an alternative genome-scale strategy to comprehensively analyze *stx* allele status and losses or gains in the strain’s phage ORF-omes.

Major virulence traits of *E. coli* O157:H7 are encoded on members of the mobilome that are usually stably integrated into the chromosome, such as the locus of enterocyte effacement (LEE) and *stx*-converting phages (Nataro and Kaper, 1998). Phages are key components of pathogenome evolution and their acquisitions are important events in the emergence of *E. coli* O157:H7 from an ancestral cell closely related to *E. coli* O55:H7 (Feng et al., 1998, 2007; Zhou et al., 2010). Moreover, analyses such as SNP typing that are limited to the core genome cannot provide information about the conferred pathogenic potential anchored in the mobilome. Our analysis of the 2006 SP outbreak exemplifies genomic heterogeneity that can be found in a single outbreak of O157:H7 in regards to mobilome (Eppinger et al., 2011b). Within the prophage pool (Hayashi et al., 2001) the *stx*-converting bacteriophages are of particular interest, as they encode a potent cytotoxin, Shiga toxin or Stx (Karmali et al., 2010) as direct mediator of EHEC O157:H7 disease (Krüger and Lucchesi, 2015). In *E. coli* O157:H7 the chromosomal backbone is highly conserved and genomic alterations chiefly relate to phage complement, plasticity, and respective integration sites (Shaikh and Tarr, 2003; Abu-Ali et al., 2009; Eppinger et al., 2011b, 2013; Smith et al., 2012; Yin et al., 2015). Three *stx* alleles, *stx1a*, *stx2a*, and *stx2c*, are found predominantly in this lineage (Scheutz et al., 2012). We used discontinuous megablast against the VirulenceFinder database to determine the toxin subtypes present in each outbreak (Joensen et al., 2014). All IH, PM, and outbreak strains carry the more potent allelic variant *stx2a* (Supplemental Table 1). In addition, all FS and TJ outbreak strains, PMs B40-1/2 and B26-1/2 and two separate IH transmission cases (B83/B84, B85/B86) carry an *stx1*-converting phage. Co-carriage of Stx2 and Stx1 can reduce Stx2a production (Serra-Moreno et al., 2008) and also attenuates end-organ toxicity of Stx2a (Donohue-Rolfe et al., 2000; Russo et al., 2016). Noteworthy, the 2006 SP outbreak associated with hypervirulence (Kulasekara et al., 2009; Abu-Ali et al., 2010) features the Stx2a/2c toxin type, with an almost complete *stx1*-converting phage occupies *yehV*. However, this atypical phage lacks *stx1*genes (Eppinger et al., 2011b). We also note that the TJ lettuce isolate TW14588 harbors two *stx2a*-converting phages integrated at *argW* and *wrbA* (Supplemental Table 1). We speculate that double *stx2a*-converting phage occupancy might also increase pathogenic potential, such as through phage dosage effects, also considering that *stx2a* is the most potent allelic subtype (Tesh et al., 1993; Tesh, 2010; Fogg et al., 2012). We note that this information cannot be gathered by PCR-based Stx-subtyping (Scheutz et al., 2012), as this approach does not determine copy number, highlighting the increased resolution obtained by WGS in regards to the pathogenic potential of the outbreak (Holmes et al., 2015). All other outbreaks except BL possess *stx2c*-converting bacteriophages. The interplay between these two *stx2*-converting phage types is not known, although both variants have been linked to HUS (Friedrich et al., 2002;



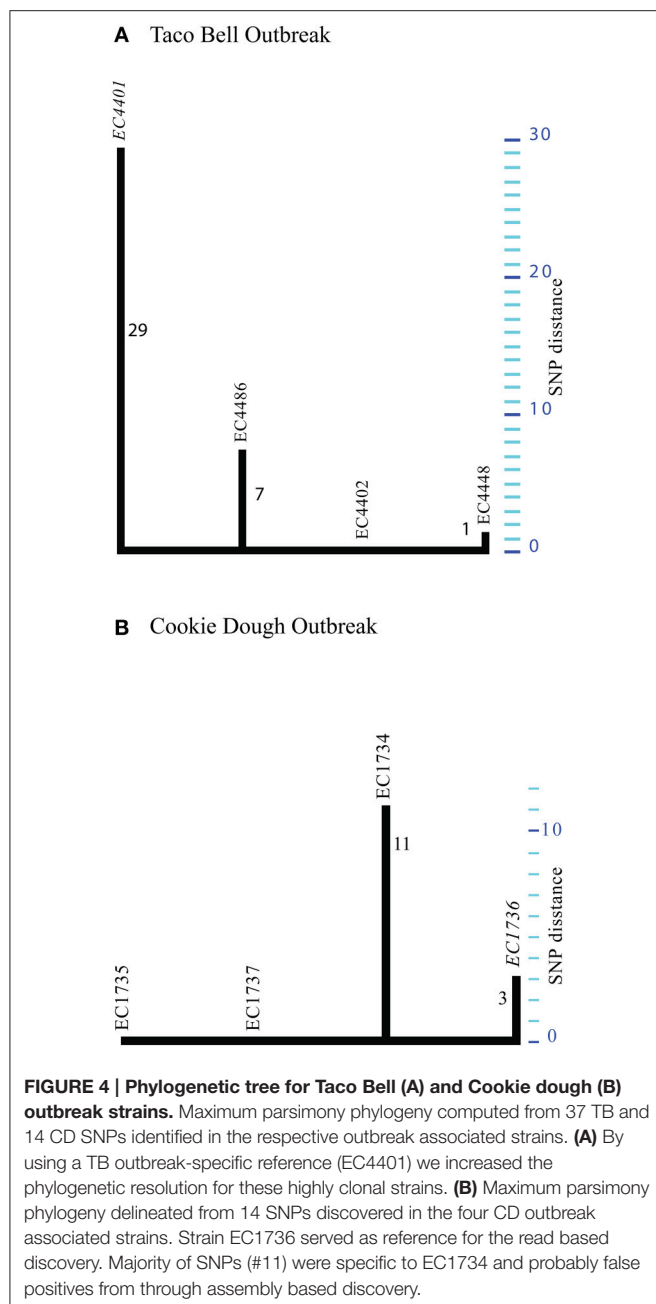
Persson et al., 2007). We observed only two variations in the *stx2* allelic profiles in the CD and TP outbreaks (Supplemental Table 1). The non-clinical CD outlier EC1738 collected in the production plant is distinguished by lack of the *stx2a* allele. TP strain EC1870 lacks the *stx2c* allele that is present in all other TP outbreak strains (Supplemental Table 1).

On average, the phage complement in the strains represents 14% of predicted coding regions in the genomes, in accordance with other studies (Asadulghani et al., 2009; Smith et al., 2012). As expected when considering the close relation between outbreak strains (Figure 2), the variability in phage-borne ORFs was low (5%) (Supplemental Figure 1; Supplemental Table 3). This small variability represents the noise caused by clustering of related proteins into centroids in the LS-BSR analysis rather than differences in phage regions. The TB associated strains had more variability (11.2%) (Supplemental Table 3), due to inclusion of the temporal outliers EC4436, EC4437, and EC4439 (Figure 6). Variome analysis highlighted phage regions that were unique to each temporal outlier group, resulting in the same clustering as in the SNP-based analysis (Figures 2, 6). The CD outbreak also had greater variability (16.5%) despite the exclusion of outlier EC1738 (Supplemental Table 3), likely attributable to differences in fragmentation of the analyzed genomes. We observed a correlation of quality of PHAST prediction with size of contigs and reduced genome fragmentation. Closed genomes and genomes with larger contigs had up to 20% more predicted phage regions that also served to increase the noise, compared to more fragmented genomes (Supplemental Table 1).

The identified phage complements of the FS isolates were highly similar. However, we found phage sequences unique to strain B103 (Supplemental Figure 2A). Discontiguous megablast of the phage region (Buhler, 2001; Ma et al.,

2002) against closed bacteriophages identified *Escherichia* phage pro483 (KR073661), originally isolated from an avian pathogenic *E. coli* DE048. This prophage was previously described in SP strains (Eppinger et al., 2011b) and a supershedder strain SS17 (Cote et al., 2015). Unlike the *yegQ* insertion in SP outbreak strains (Eppinger et al., 2011b), this phage disrupts the colicin immunity protein (WP_001303895) in strain B103. Using the phage pro483 (KR073661) as a genomic anchor for the B103 draft contigs we recovered 86% of the phage genome with 97% identity. We further identified a 12 bp (ACCAATAACTGA) repeat at both ends of the phage borders, indicative of the phage integration mechanism (Campbell, 1992). The SP outbreak strain EC4115 however features an 18 bp repeat (Eppinger et al., 2011b). The genomic architecture is syntenic and largely conserved throughout the phage genome, except for insertion or deletion introduced by an exonuclease (ECH74115_RS15445) and a hypothetical protein (ECH74115_RS15450) only present in EC4115 (Supplemental Figure 3).

Acquisition or loss of phages secondary to recombination events during the course of an outbreak creates interstrain plasticity. Thus, analysis of a single archetypical outbreak strain might underestimate the mobilome and core chromosome plasticity (Eppinger and Cebula, 2015). Comprehensive analyses did not reveal significant differences in phage content of the BL and FF outbreak clusters (Figure 2) to further distinguish these clonal outbreaks featuring only one and four SNPs per outbreak cluster, respectively (Supplemental Figures 2B,C). In contrast, the TP outbreak strains displayed a higher degree of mobilome plasticity (Supplemental Figure 2D), in line with the higher number of predicted SNPs (#98) for this outbreak (Figure 2, Supplemental Dataset 1).



Plasmid Prevalence in Clinical *E. coli* O157:H7 Strains

The *E. coli* O157:H7 lineage is distinguished from other serotypes by the presence of the large virulence plasmid pO157 (Burland et al., 1998). For this serotype, additional plasmids have been occasionally characterized at sequence level (Makino et al., 1998; Eppinger et al., 2011b) or by plasmid profiling (Ostroff et al., 1989; Meng et al., 1995). To facilitate plasmid discovery and survey we reassembled the genomes with SPAdes (Bankevich et al., 2012). Even though deposited genomes from 454 and Illumina Celera hybrid assemblies (Denisov et al., 2008) had fewer contigs compared to SPAdes assemblies from Illumina

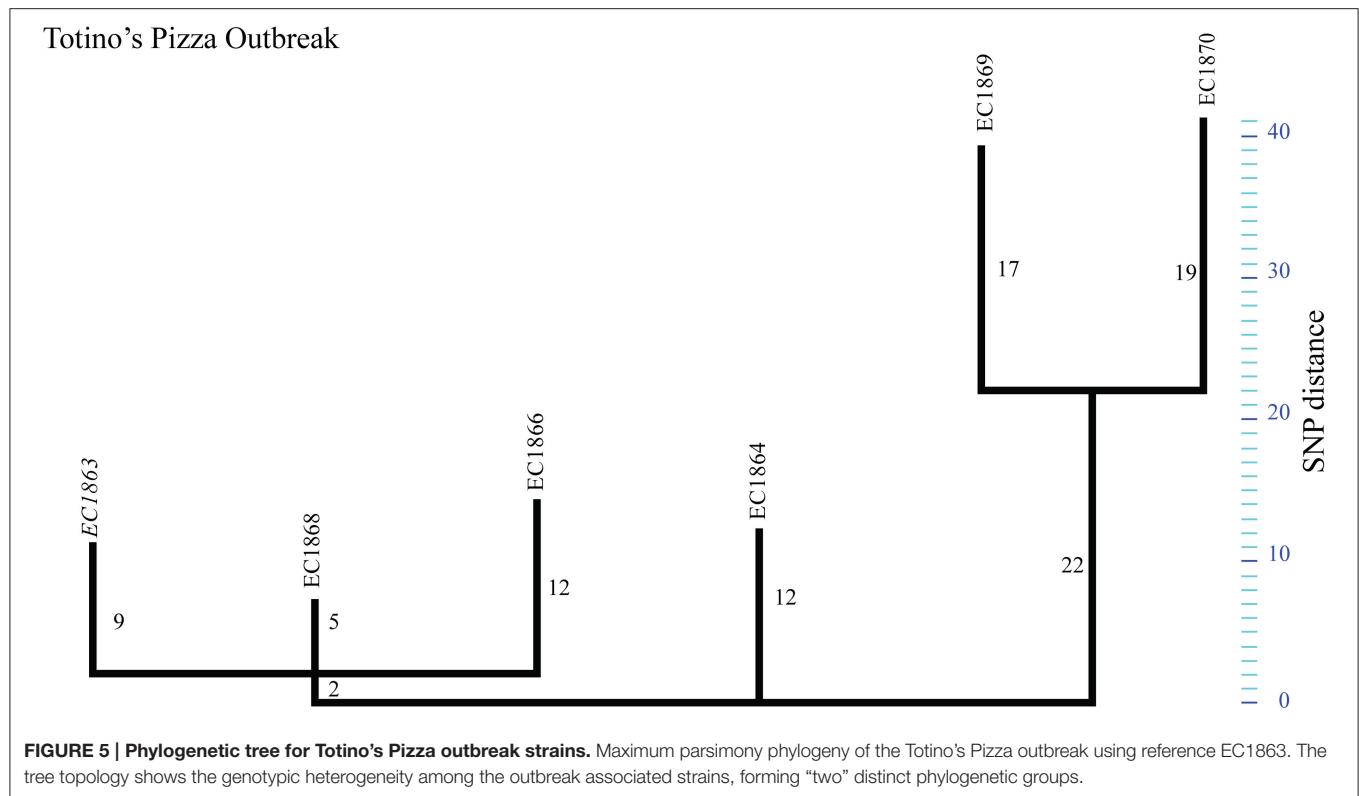
reads only (Supplemental Table 1), reassembly typically produced longer contigs, in particular for plasmid-originating regions. If Illumina reads only were processed, the SPAdes assemblies clearly outperformed NCBI deposited Velvet assemblies in regards to sensitivity for plasmid prediction (Supplemental Table 1). We queried plasmid sequences against the NCBI nr plasmid database using discontinuous megablast (Buhler, 2001; Ma et al., 2002).

Using this approach we discovered five plasmids at the sequence level (p78, p34, p55, p63, and p39) that have not been previously described in deposited *E. coli* O157:H7 genomes. Among these is a homolog of a 37 kb conjugal transfer pEC4115, referred to as p36, originally described in the SP outbreak strains (Eppinger et al., 2011b). We found the TB and TP outbreaks to be most diverse in regards to plasmid carriage (Figure 2). The TB associated strains contained three distinct plasmid profiles, which correlated with the clustering from the core genome SNP discovery (Figure 2).

Plasmid p78, the largest plasmid, shows homology to the conjugative IncI1 group *E. coli* plasmid pC49-108 and *Salmonella enterica* plasmids (Fricke et al., 2011; Kröger et al., 2012; Wang et al., 2014a). p78 varied in length, from 78 to 88 kb in clade 8 strains (Figure 7). The related plasmid pC49-108 carries multiple antibiotic resistance genes (Wang et al., 2014a), including a beta-lactamase (*bla*_{CTX-M-1}) (Wang et al., 2014a), dihydrofolate reductase (*dfrA17*) and aminoglycoside adenyltransferase (*aadA5*) found both adjacent to a class 1 integron (Wang et al., 2014a). In similarity to the *bla*_{CTX-M-1} located next to a mobile element (ISECp1), we found another class C beta-lactamase gene in *S. enterica* CVM 22462, again found next to a mobile transposase locus. We speculate that colocalization to mobile elements might affect locus stability and explains the scattered prevalence of these resistances in the plasmid homologs (Wang et al., 2014b) (Figure 7).

Resistance to antibiotics has been observed in *E. coli* O157:H7, but the genetic basis remains largely unknown (Meng et al., 1998). We previously linked multi drug resistance (MDR), a rare occurrence in *E. coli* O157:H7, to phage-borne antibiotic resistance loci (Eppinger et al., 2011a). Strain EC4402, part of the core TB outbreak cluster, was identified as a MDR isolate (Figure 2). This strain displays elevated MICs for several cephalosporins and aminoglycosides, sulfisoxazole and nalidixic acid (quinolone). However, our *in silico* analysis with ResFinder (Kleinheinz et al., 2014) did not reveal any potential underlying resistance loci. Here we note that resistance phenotypes can be conferred by loci not previously linked to antibiotic resistance (Gibson et al., 2016). We speculate that the resistance loci might have been either lost from the original p78, or were an integral part of other MDR plasmids lost during laboratory cultivation prior to the sequencing of EC4402. Alternatively, the antibiotic resistance might be conferred by yet unknown loci not represented in queried resistance databases.

Plasmid p36 was highly homologous to other conjugal transfer plasmids, such as *S. enterica* plasmid pCFSAN000111_01 (NZ_CP007599) (Timme et al., 2012) and pEC4115 (Eppinger et al., 2011b) (Supplemental Figure 4). While p78 was found solely in clade 8, p36 seems to be more widespread (Cote et al., 2015), and present in non-O157:H7 *E. coli* serotypes.



Co-carriage of a p78-p36 combination was also found in clade 8 strains K4405 and K4406 (**Figure 7**, Supplemental Figure 4). The TB outlier strain EC4439 lacks both p78 and p36, but carries a p55 plasmid with high homology to *Klebsiella pneumoniae* pDMC1097-77.775 kb (87% coverage, 99% identity) (Wright et al., 2014) (Supplemental Figure 5). This IncI2 group plasmid carries multiple resistances, which are absent in the *E. coli* plasmid homologs (Supplemental Figure 5). Interestingly, this plasmid is also present in IH strain B86, but absent from strain B85, either because of independent acquisition or secondary loss in B85, respectively (**Figure 2**). Our findings on plasmid prevalence are in accordance with those of Dallman et al. who showed that epidemiologically linked strains can vary largely in their plasmid inventory (Holmes et al., 2015).

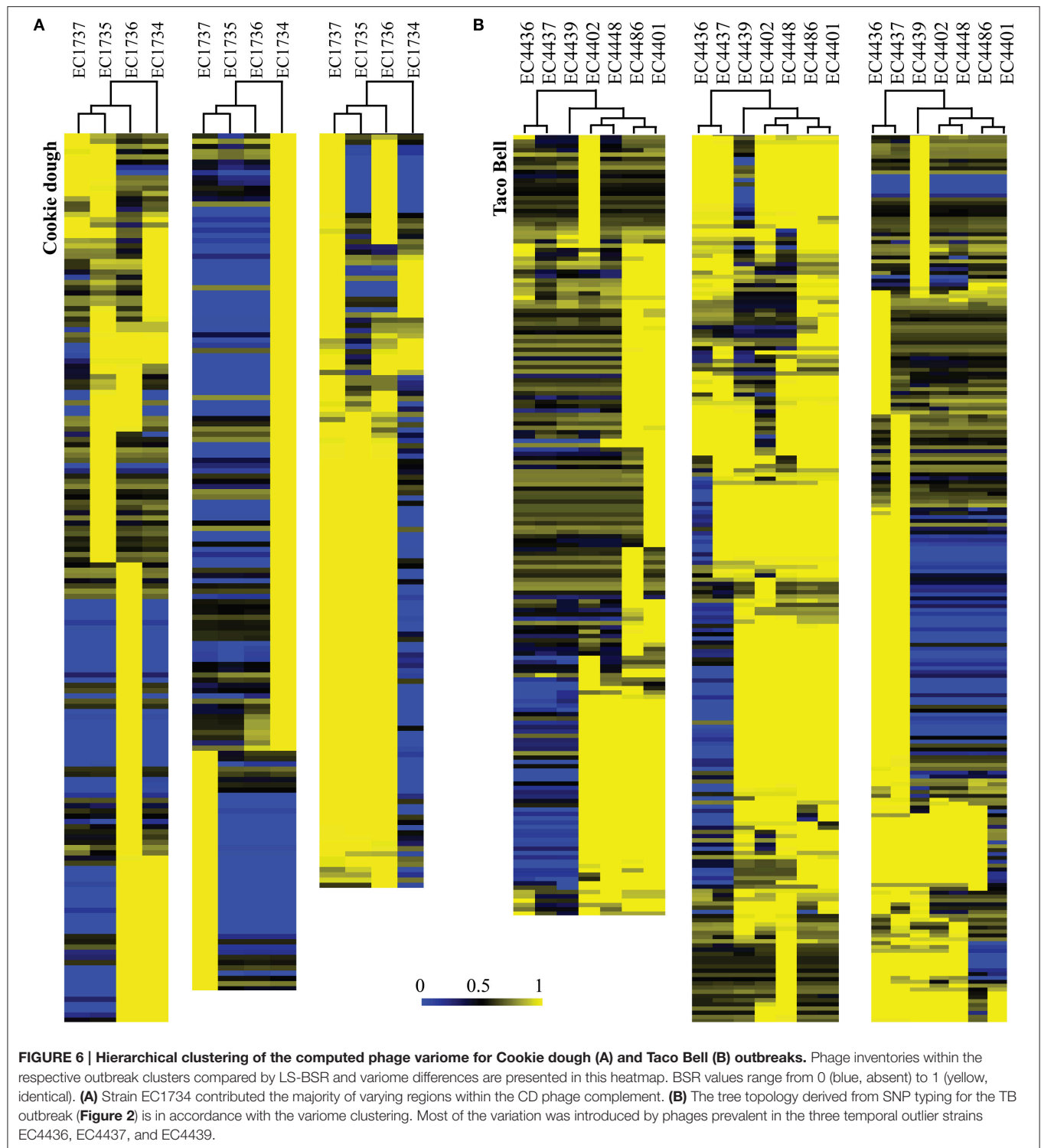
The IH strains B89 and B90 harbor p34 (Supplemental Figure 6), which is related to *E. coli* pVR50, an F-like conjugative MDR plasmid (Beatson et al., 2015). While the overall plasmid backbone is conserved, p34 lacks any resistance loci (Supplemental Figure 6). The TP strains also possess strain-specific plasmids: p36 in EC1870, p63 in EC1863, and p38 in EC1868 (**Figure 2**). Plasmid p63 has partial homology to pO26-Vir, an IncK group plasmid, a mosaic of multiple plasmids (Fratamico et al., 2011). In 1863 (p63) we found homologous loci for conjugal transfer and type IV pili (Supplemental Figure 7), which have been implicated in cell adherence and biofilm formation (Dudley et al., 2006), and notably, these phenotypes are strain-dependent in *E. coli* O157:H7 (Vogeleer et al., 2014). A 39 kb plasmid fragment in EC1868 (p39) was found to be

homologous to a 87 kb IncFII plasmid from *E. coli* (pGUE-NDM) (Bonnin et al., 2012) (**Figure 2**).

The observed variability in plasmid type and prevalence in the individual strains clearly highlights genomic plasticity that exists even among closely related isolates of the same origin and can be utilized for strain attribution (Eppinger et al., 2011b). The identified heterogeneity between the mobilome of outbreak strains stresses the importance of studying a number of isolates from the same outbreak instead of using archetypal outbreak strains, which as shown might not fully reflect the plasticity in the outbreak population (Eppinger and Cebula, 2015). Interestingly, all the above described *E. coli* O157:H7 plasmids lack antibiotic resistance loci, even though our plasmid survey found widespread resistances among homologous plasmids in other serotypes and species.

CONCLUSION

While some of these clinical isolates have been studied previously using molecular epidemiology techniques (Samadpour et al., 2002), we have for the first time applied whole genomics epidemiology approaches (Eppinger et al., 2011b). Through these high resolution methods we established a detailed understanding of the genomic heterogeneity found among the studied *E. coli* O157:H7 outbreak populations from the U.S. The gathered phylogenomic data were critical to define the genetic relatedness of individual strains in the context of outbreak etiology and phylogenetic positions in the broader model of *E. coli* O157:H7 evolution and epidemiology. In this



study, we detected previously unnoted polymorphic genome features in the core and mobile genome, such as an array of new plasmids not previously associated with this lineage. The cataloged polymorphic signatures aided in strain attribution and allowed us to precisely define the outbreak boundaries. This

allowed us to discern the distinct phylogenetic boundaries of the studied EHEC strains when placed into a larger phylogenomic framework of *E. coli* O157:H7 from North America (**Figure 2**) assessing both core and mobilome (**Figures 6, 7**) (Eppinger and Cebula, 2015). The developed WGS typing approach

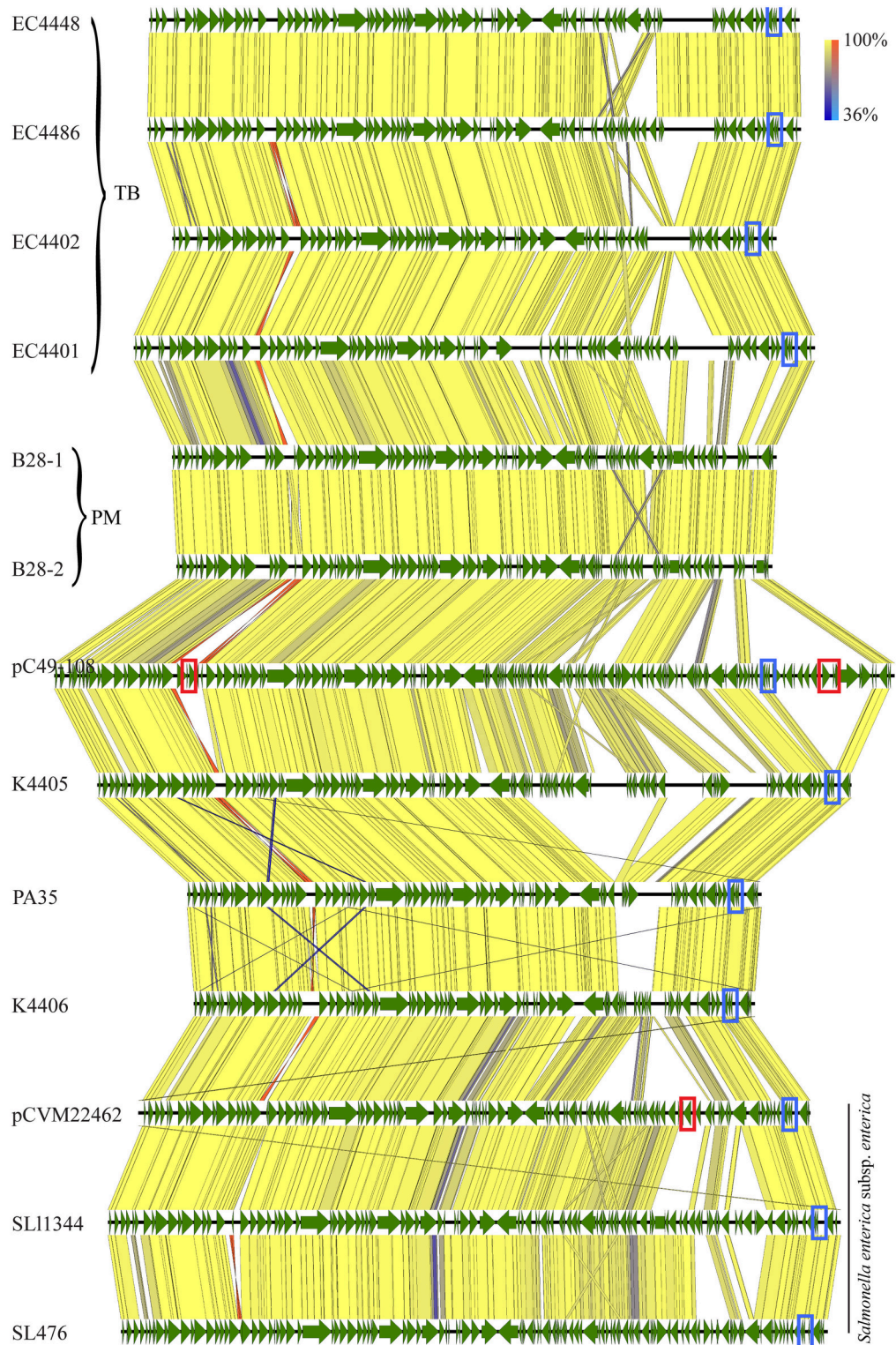


FIGURE 7 | Alignment of plasmid p78 and homologs. Plasmid architecture and gene inventories were compared by tblastx, and respective annotations were mapped in Geneious vR9 (CDS, green). The plasmid architecture was highly conserved with a high identity level throughout the entire length of the plasmid [100% identity yellow (inverted fragment, orange), 36% identity blue (inverted fragment light blue)]. Plasmid p78 homologs were widespread, such as in TB outbreak associated strains, B28 plate mates, other *E. coli* O157:H7, and *S. enterica*. We found a locus for a RelB/E tox/antitoxin system present in all plasmids, with the exception of PM strains B28 (blue box). Resistance loci are highlighted in red boxes.

TABLE 6 | Proposed criteria and practices for SNP-based epidemiological outbreak inclusion or exclusion.

Ideal case	
PRIMARY CRITERIA	
Highly credible exposure	Point source (clustered in time and space), pathogen isolated from incriminated vehicle
High quality sequencing	Contig numbers should be comparable to previous assemblies performed with the same assembly method, minimum coverage required will depend on technology used
Mobilome exclusion	Data focus on the most immutable part of the genome, most commonly the genome backbone
SNP validation	Second method, PCR confirmation or re-sequencing to validate mutations in discriminatory SNPs
SECONDARY CRITERIA	
Collection of multiple isolates from cases for accurate attribution	The larger the sample that demonstrates homogeneity, the greater the likelihood of a common source
	If a common PFGE/MLVA type is present in the same region, confirmation that allelic differences exist between outbreak strain and background non-outbreak strains
SNP calling	Reference and outbreak based SNP comparisons
	Allelic frequency >=0.9
	Base and mapping quality control
Novel SNP	If a tolerance is set at >0 for SNPs as a cut point for assigning isolates as being from the same source, a SNP that is not in the database (i.e., apparently <i>de novo</i>) would be given more credence than one that has been described previously, and which probably represents a different lineage or located in a known polymorphic gene, such as <i>rpoS</i>
QUESTIONABLE CRITERIA	
Variances in the complementary mobilome, such as presence of plasmids and phages	Loci of likely mobile origin are not reliable as a differentiating trait for epidemiologic purposes
BEST PRACTICES IN REPORTING RESULTS	
Report exclusion criteria	A list of loci and regions that were excluded from SNP discovery
Provide list of SNP loci and alleles	Provide information on location of SNP, product, resulting codon change
Provide reads	Deposition of sequences and strain associated metadata in public repositories

provided us with the necessary resolution power to study the individual dynamics in highly clonal outbreaks (Morelli et al., 2010; Eppinger et al., 2011b, 2014; Hasan et al., 2012; Berenger et al., 2015; Holmes et al., 2015; Jenkins et al., 2015).

While the majority of outbreaks were caused by pathogens that form tight clonal clusters, one outbreak (“Totino’s Pizza”) was associated with isolates showing considerable genomic heterogeneity (Figures 2, 5). Apparent SNPs in other outbreaks are associated with a paucity of reads for quality control, falsely increasing the diversity among the outbreak isolates. Since outbreaks can have high economic impacts, such as nationwide recalls of contaminated product, multiple samples from the same outbreak should be concomitantly sequenced instead of using archetypal outbreak strains to provide strong evidence for inclusion or exclusion, strain and source attribution (Eppinger et al., 2010, 2014; Morelli et al., 2010; Hasan et al., 2012). Additionally, these high resolution approaches allow for the discovery of emerging pathotypes, and, potentially, to better assess the pathogenic potential of individual bacterial clones (Berenger et al., 2015; Klemm and Dougan, 2016). Expanding these sequence-based analyses to the publicly available EHEC sequence pool will improve public health response in the event of an outbreak allowing timely and informed countermeasures. Canonical SNPs can be implemented in efficient typing assays offering robust phylogenetic signals for outbreak exclusion/inclusion that surpass classical technologies

(Riordan et al., 2008; Elhadidy et al., 2015; Rusconi and Eppinger, 2016).

Our study strongly endorses that quality of SNPs and choice of an appropriate reference strain in WGST approaches are equally critical to achieve phylogenetic resolution and accuracy (Table 6). Here we also demonstrate that in order to avoid type 2 error of attribution, the quality of SNP data obtained from WGS approaches is crucial (Table 6). For read-based discovery approaches we would like to emphasize the importance of SRA data availability, which is not only foundational to determine coverage and quality of detected SNP positions, but also to optimize assembly quality should assemblers with improved algorithms become available (Supplemental Table 1). SNP discovery with an appropriate outbreak-specific reference strain is critical for reference based WGS typing. To fully assess the genomic plasticity, the reference should be phylogenetically related and not too distant to the strains of interest, as evidenced by the resolution power gained using a within outbreak reference (Figures 2, 4, 5). By extending our analysis to the mobilome, we detected plasticity among clonal strains in phage and plasmid content describing novel plasmids not previously associated with *E. coli* O157:H7. We also would like to stress the importance of publicly available strain associated clinical, environmental, and epidemiological metadata concomitantly to the genomic data as prerequisite for informed source attribution (Table 6) (Eppinger and Cebula, 2015). We anticipate that NGS long-read technology, such as contemporary SMRT technology (English et al., 2012), or

other platforms under development (Feng et al., 2015; Rhoads and Au, 2015) will tremendously benefit WGS typing strategies as it pertains to the highly homogenous *E. coli* O157:H7 lineage (Zhang et al., 2006, 2007; Eppinger et al., 2011b). In particular, long-read technologies will produce (near) closed genomes and thus allow to accurately determine the *stx*-virulence status by defining not only *stx* allele type, but also *stx*-converting phage combination, plasticity, and location, all factors that have been associated with alterations in Stx-production as direct mediator of EHEC disease (Ogura et al., 2015; Toro et al., 2015; Yin et al., 2015).

Our data provide insight into the maximum number of permissible SNPs two strains can have and still designate them of the same origin. In prior work, we found no SNPs between 24 isolates of the same point-source cluster, focusing on backbone ORFs (Turabelidze et al., 2013). Dallman et al. and others tolerated up to 4 SNPs in the core genome before assigning two isolates to different sources (Underwood et al., 2013; Joensen et al., 2014; Dallman et al., 2015; Holmes et al., 2015). We found one *bona fide* SNP in the course of a single point-source, short-term outbreak. Since no gold standards have yet been accepted for *E. coli* O157:H7 WGS typing we propose the following criteria (Table 6) for inclusion (presumably of same source) vs. exclusion (presumably of different source) of investigated isolates: (i) High-quality whole genome sequence fortified with extensive epidemiological outbreak data, (ii) genome-scale SNP discovery based on high quality sequencing with reference, (iii) exclusion of mobilome and repeats (to reduce epidemiological noise), followed by (iv) PCR-confirmation of eventual SNPs for definitive in-/exclusion, and (v) mobilome discovery which can significantly contribute to the genomic plasticity. Moreover, for cases that are quite dispersed in time and space, there should be greater stringency in assigning “like” status to two strains that are even differentiated by a single SNP. When outbreaks occur, there are often large product liability issues at stake, and considerable obligation on disease control authorities to identify such clusters and molecular typing serves an increasingly important role. Therefore, diligence should be exercised in choice of sequence-based typing protocols, and in their analysis.

Finally, while we eagerly anticipate the introduction of sequence-based pathogen typing as a public health and disease prevention tool (Sadiq et al., 2014; Eppinger and Cebula, 2015), we share the concern of Osterholm (2015), who stresses that this powerful technology be employed as an adjunct to, and not a replacement for, case interviewing (descriptive epidemiology) and environmental investigations. Also, we are entering an era of non-culture diagnosis of enteric infections, including those caused by *E. coli* O157:H7 (Schatz and Phillippy, 2012; Klemm and Dougan, 2016). The high resolution data presented in this article would not have been possible without classic diagnostic microbiology laboratory recovery of the pathogen of interest. We hope that resources will be devoted to recovering these agents from submitted specimens, so as to complement case investigation by local healthy jurisdictions.

ACCESSION NUMBER

The sequence data sets analyzed in this study have been retrieved from the short read archive (SRA) and whole genome shotgun repository at NCBI. Accession numbers for the genomes are provided in Supplemental Table 1.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: ME. Analyzed the data: BR, FS, SK, MM, PT, ME. Contributed reagents/materials/analysis tools: MM, PT, ME. Wrote the paper: BR, SK, MM, PT, ME.

FUNDING

The study was supported by the National Institute of Allergy and Infectious Diseases, National Institute of Health, Department of Health and Human Services under contract SC2AI120941, the US Department of Homeland Security under contract 2014-ST-062-000058, the Department of Biology, the South Texas Center for Emerging Infectious Diseases (STCEID) at the University of Texas at San Antonio, and the High Performance Computing Center (HPC) under contract 2G12RR013646-12. Strain collection and archiving was funded by R01DK52081 and P30DK052574 for the Biobank Core. BR was supported in part by the Swiss National Science Foundation Early Postdoc Mobility Fellowship (P2LAP3-151770). FS was supported in part by the South Texas Center for Emerging Infectious Diseases (STCEID) and through an UTSA Teaching Fellowship (UTF).

ACKNOWLEDGMENTS

This work received computational support from Computational System Biology Core, funded by the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health. We would like to thank Armando L. Rodriguez for technical support and maintenance of the Galaxy platform and Dr. Anna Allue-Guardia and Heidi Gildersleeve for critically reading the manuscript. Further we are grateful to Dr. Nurmohammad Shaikh for assistance with SNP PCR confirmation. We are also grateful to the diagnostic microbiologists, whose diligence in identifying the isolates on the original agar plates were necessary to the subsequent sequences that underlie our analysis. We are further indebted to the various outbreak investigation teams that identified the outbreaks described in this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00985>

REFERENCES

- Abu-Ali, G. S., Lacher, D. W., Wick, L. M., Qi, W., and Whittam, T. S. (2009). Genomic diversity of pathogenic *Escherichia coli* of the EHEC 2 clonal complex. *BMC Genomics* 10:296. doi: 10.1186/1471-2164-10-296
- Abu-Ali, G. S., Ouellette, L. M., Henderson, S. T., Lacher, D. W., Riordan, J. T., Whittam, T. S., et al. (2010). Increased adherence and expression of virulence genes in a lineage of *Escherichia coli* O157:H7 commonly associated with human infections. *PLoS ONE* 5:e10167. doi: 10.1371/journal.pone.0010167
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Asadulghani, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., Iguchi, A., et al. (2009). The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.* 5:e1000408. doi: 10.1371/journal.ppat.1000408
- 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
- Beaton, S. A., Ben Zakour, N. L., Totsika, M., Forde, B. M., Watts, R. E., Mabbett, A. N., et al. (2015). Molecular analysis of asymptomatic bacteriuria *Escherichia coli* strain VR50 reveals adaptation to the urinary tract by gene acquisition. *Infect. Immun.* 83, 1749–1764. doi: 10.1128/IAI.02810-14
- Bell, B. P., Goldoft, M., Griffin, P. M., Davis, M. A., Gordon, D. C., Tarr, P. I., et al. (1994). A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. *JAMA* 272, 1349–1353. doi: 10.1001/jama.1994.03520170059036
- Bentley, S. D., and Parkhill, J. (2015). Genomic perspectives on the evolution and spread of bacterial pathogens. *Proc. Biol. Sci.* 282:20150488. doi: 10.1098/rspb.2015.0488
- Berenger, B. M., Berry, C., Peterson, T., Fach, P., Delannoy, S., Li, V., et al. (2015). The utility of multiple molecular methods including whole genome sequencing as tools to differentiate *Escherichia coli* O157:H7 outbreaks. *Euro Surveill.* 20:30073. doi: 10.2807/1560-7917.ES.2015.20.47.30073
- Besser, T. E., Shaikh, N., Holt, N. J., Tarr, P. I., Konkel, M. E., Malik-Kale, P., et al. (2007). Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157:H7 isolates from cattle than in those from humans. *Appl. Environ. Microbiol.* 73, 671–679. doi: 10.1128/AEM.01035-06
- Bonnin, R. A., Poirel, L., Carattoli, A., and Nordmann, P. (2012). Characterization of an IncFII plasmid encoding NDM-1 from *Escherichia coli* ST131. *PLoS ONE* 7:e34752. doi: 10.1371/journal.pone.0034752
- Buhler, J. (2001). Efficient large-scale sequence comparison by locality-sensitive hashing. *Bioinformatics* 17, 419–428. doi: 10.1093/bioinformatics/17.5.419
- Burland, V., Shao, Y., Perna, N. T., Plunkett, G., Sofia, H. J., and Blattner, F. R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* 26, 4196–4204. doi: 10.1093/nar/26.18.4196
- Campbell, A. M. (1992). Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* 174, 7495–7499.
- Cornick, N. A., Jelacic, S., Ciol, M. A., and Tarr, P. I. (2002). *Escherichia coli* O157:H7 infections: discordance between filterable fecal shiga toxin and disease outcome. *J. Infect. Dis.* 186, 57–63. doi: 10.1086/341295
- Cote, R., Katani, R., Moreau, M. R., Kudva, I. T., Arthur, T. M., Debroy, C., et al. (2015). Comparative analysis of super-shedder strains of *Escherichia coli* O157:H7 reveals distinctive genomic features and a strongly aggregative adherent phenotype on bovine rectoanal junction squamous epithelial cells. *PLoS ONE* 10:e0116743. doi: 10.1371/journal.pone.0116743
- Dallman, T. J., Byrne, L., Ashton, P. M., Cowley, L. A., Perry, N. T., Adak, G., et al. (2015). Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin. Infect. Dis.* 61, 305–312. doi: 10.1093/cid/civ318
- 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
- Davis, M. A., Hancock, D. D., Besser, T. E., and Call, D. R. (2003). Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of *Escherichia coli* O157:H7. *J. Clin. Microbiol.* 41, 1843–1849. doi: 10.1128/JCM.41.5.1843-1849.2003
- Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics Chapter 10:Unit* 10.13. doi: 10.1002/0471250953.bi1003s00
- den Bakker, H. C., Allard, M. W., Bopp, D., Brown, E. W., Fontana, J., Iqbal, Z., et al. (2014). Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar enteritidis. *Emerg. Infect. Dis.* 20, 1306–1314. doi: 10.3201/eid2008.131399
- Denisov, G., Walenz, B., Halpern, A. L., Miller, J., Axelrod, N., Levy, S., et al. (2008). Consensus generation and variant detection by Celera assembler. *Bioinformatics* 24, 1035–1040. doi: 10.1093/bioinformatics/btn074
- Donohue-Rolfe, A., Kondova, I., Oswald, S., Hutto, D., and Tzipori, S. (2000). *Escherichia coli* O157:H7 strains that express Shiga toxin (Stx) 2 alone are more neurotropic for gnotobiotic piglets than are isolates producing only Stx1 or both Stx1 and Stx2. *J. Infect. Dis.* 181, 1825–1829. doi: 10.1086/315421
- Dudley, E. G., Abe, C., Ghigo, J. M., Latour-Lambert, P., Hormazabal, J. C., and Nataro, J. P. (2006). An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative *Escherichia coli* strain C1096 to cultured cells and abiotic surfaces. *Infect. Immun.* 74, 2102–2114. doi: 10.1128/IAI.74.4.2102-2114.2006
- Elhadidy, M., Elkhatib, W. F., Piérard, D., De Reu, K., and Heyndrickx, M. (2015). Model-based clustering of *Escherichia coli* O157:H7 genotypes and their potential association with clinical outcome in human infections. *Diagn. Microbiol. Infect. Dis.* 83, 198–202. doi: 10.1016/j.diagmicrobio.2015.06.016
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., et al. (2012). Mind the gap: upgrading genomes with Pacific biosciences RS long-read sequencing technology. *PLoS ONE* 7:e47768. doi: 10.1371/journal.pone.0047768
- Eppinger, M., and Cebula, T. A. (2015). Future perspectives, applications and challenges of genomic epidemiology studies for food-borne pathogens: a case study of Enterohemorrhagic *Escherichia coli* (EHEC) of the O157:H7 serotype. *Gut Microbes* 6, 194–201. doi: 10.4161/19490976.2014.969979
- Eppinger, M., Daugherty, S., Agrawal, S., Galens, K., Sengamalay, N., Sadzewicz, L., et al. (2013). Whole-genome draft sequences of 26 enterohemorrhagic *Escherichia coli* O157:H7 strains. *Genome Announc.* 1:e00134-00112. doi: 10.1128/genomeA.00134-12
- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011a). Genome Signatures of *Escherichia coli* O157:H7 Isolates from the bovine host reservoir. *Appl. Environ. Microbiol.* 77, 2916–2925. doi: 10.1128/AEM.02554-10
- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011b). Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20142–20147. doi: 10.1073/pnas.1107176108
- Eppinger, M., Pearson, T., Koenig, S. S., Pearson, O., Hicks, N., Agrawal, S., et al. (2014). Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *MBio* 5:e01721. doi: 10.1128/mBio.01721-14
- Eppinger, M., Worsham, P. L., Nikolich, M. P., Riley, D. R., Sebastian, Y., Mou, S., et al. (2010). Genome sequence of the deep-rooted yersinia pestis strain angola reveals new insights into the evolution and pangenome of the plague bacterium. *J. Bacteriol.* 192, 1685–1699. doi: 10.1128/JB.01518-09
- Feng, P. C., Monday, S. R., Lacher, D. W., Allison, L., Siitonen, A., Keys, C., et al. (2007). Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. *Emerg. Infect. Dis.* 13, 1701–1706. doi: 10.3201/eid1311.070381
- Feng, P., Lampel, K. A., Karch, H., and Whittam, T. S. (1998). Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J. Infect. Dis.* 177, 1750–1753. doi: 10.1086/517438
- Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* 13, 4–16. doi: 10.1016/j.gpb.2015.01.009
- Finn, R. D., Cogill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Fogg, P. C., Saunders, J. R., McCarthy, A. J., and Allison, H. E. (2012). Cumulative effect of prophage burden on Shiga toxin production in *Escherichia coli*. *Microbiology* 158, 488–497. doi: 10.1099/mic.0.054981-0

- Fratamico, P. M., Yan, X., Caprioli, A., Esposito, G., Needleman, D. S., Pepe, T., et al. (2011). The complete DNA sequence and analysis of the virulence plasmid and of five additional plasmids carried by Shiga toxin-producing *Escherichia coli* O26:H11 strain H30. *Int. J. Med. Microbiol.* 301, 192–203. doi: 10.1016/j.ijmm.2010.09.002
- Freedman, S. B., Xie, J., Neufeld, M. S., Hamilton, W. L., Hartling, L., Tarr, P. I., et al. (2016). Shiga toxin-producing *Escherichia coli* Infection, antibiotics, and risk of developing hemolytic uremic syndrome: a meta-analysis. *Clin. Infect. Dis.* 62, 1251–1258. doi: 10.1093/cid/ciw099
- Fricke, W. F., Mammel, M. K., McDermott, P. F., Tartera, C., White, D. G., Leclerc, J. E., et al. (2011). Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. *J. Bacteriol.* 193, 3556–3568. doi: 10.1128/JB.00297-11
- Friedrich, A. W., Bielaszewska, M., Zhang, W. L., Pulz, M., Kuczius, T., Ammon, A., et al. (2002). *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J. Infect. Dis.* 185, 74–84. doi: 10.1086/338115
- Garrison, E., and Marth, G. (2012). *Haplotype-Based Variant Detection from Short-Read Sequencing*. ArXiv e-prints [Online], 1207. Available online at: <http://adsabs.harvard.edu/abs/2012arXiv1207.3907G> (Accessed July 1, 2012).
- Gibson, M. K., Wang, B., Ahmadi, S., Burnham, C.-A. D., Tarr, P. I., Dantas, G., et al. (2016). Developmental dynamics of the preterm infant gut microbiota and antibiotic resistance. *Nat. Microbiol.* 1:16024. doi: 10.1038/nmicrobiol.2016.24
- Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A. Jr., and Hewlett, E. L. (2015). Whole-genome sequencing in outbreak analysis. *Clin. Microbiol. Rev.* 28, 541–563. doi: 10.1128/CMR.00075-13
- Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86
- Goldwater, P. N., and Bettelheim, K. A. (2012). Treatment of enterohemorrhagic *Escherichia coli* (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med.* 10:12. doi: 10.1186/1741-7015-10-12
- Griffing, S. M., MacCannell, D. R., Schmidtke, A. J., Freeman, M. M., Hyytia-Trees, E., Gerner-Smidt, P., et al. (2015). Canonical single nucleotide polymorphisms (SNPs) for high-resolution subtyping of shiga-toxin producing *Escherichia coli* (STEC) O157:H7. *PLoS ONE* 10:e0131967. doi: 10.1371/journal.pone.0131967
- Hasan, N. A., Choi, S. Y., Eppinger, M., Clark, P. W., Chen, A., Alam, M., et al. (2012). Genomic diversity of 2010 Haitian cholera outbreak strains. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2010–E2017. doi: 10.1073/pnas.1207359109
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., et al. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22. doi: 10.1093/dnares/8.1.11
- Holmes, A., Allison, L., Ward, M., Dallman, T. J., Clark, R., Fawkes, A., et al. (2015). Utility of whole-genome sequencing of *Escherichia coli* O157 for outbreak detection and epidemiological surveillance. *J. Clin. Microbiol.* 53, 3565–3573. doi: 10.1128/JCM.01066-15
- 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
- Institute, C. A. L. S. (2013). *Performance Standards for Antimicrobial Susceptibility Testing: 23rd Informational Supplement*. Wayne, PA: CLSI.
- Jenkins, C., Dallman, T. J., Launders, N., Willis, C., Byrne, L., Jorgensen, F., et al. (2015). Public health investigation of two outbreaks of shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. *Appl. Environ. Microbiol.* 81, 3946–3952. doi: 10.1128/AEM.04188-14
- Joensen, K. G., Scheut, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Karmali, M. A., Gannon, V., and Sargeant, J. M. (2010). Verocytotoxin-producing *Escherichia coli* (VTEC). *Vet. Microbiol.* 140, 360–370. doi: 10.1016/j.vetmic.2009.04.011
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kleinheinz, K. A., Joensen, K. G., and Larsen, M. V. (2014). Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* 4:e27943. doi: 10.4161/bact.27943
- Klemm, E., and Dougan, G. (2016). Advances in understanding bacterial pathogenesis gained from whole-genome sequencing and phylogenetics. *Cell Host Microbe* 19, 599–610. doi: 10.1016/j.chom.2016.04.015
- Kotewicz, M. L., Mammel, M. K., LeClerc, J. E., and Cebula, T. A. (2008). Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* 154, 3518–3528. doi: 10.1099/mic.0.2008/019026-0
- Kröger, C., Dillon, S. C., Cameron, A. D., Papenfort, K., Sivasankaran, S. K., Hokamp, K., et al. (2012). The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1277–E1286. doi: 10.1073/pnas.1201061109
- Krüger, A., and Lucchesi, P. M. (2015). Shiga toxins and stx phages: highly diverse entities. *Microbiology* 161, 451–462. doi: 10.1099/mic.0.000003
- Kulasekara, B. R., Jacobs, M., Zhou, Y., Wu, Z., Sims, E., Saenphimmachak, C., et al. (2009). Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect. Immun.* 77, 3713–3721. doi: 10.1128/IAI.00198-09
- Kyle, J. L., Cummings, C. A., Parker, C. T., Quiñones, B., Vatta, P., Newton, E., et al. (2012). *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. *J. Bacteriol.* 194, 1885–1896. doi: 10.1128/JB.00120-12
- Lambert, D., Carrillo, C. D., Koziol, A. G., Manninger, P., and Blais, B. W. (2015). GeneSippr: a rapid whole-genome approach for the identification and characterization of foodborne pathogens such as priority Shiga toxigenic *Escherichia coli*. *PLoS ONE* 10:e0122928. doi: 10.1371/journal.pone.0122928
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Latif, H., Li, H. J., Charusanti, P., Palsson, B. Ø., and Aziz, R. K. (2014). A Gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Genome Announc.* 2:e00821-14. doi: 10.1128/genomeA.00821-14
- Latreille, P., Norton, S., Goldman, B. S., Henkhaus, J., Miller, N., Barbazuk, B., et al. (2007). Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genomics* 8:321. doi: 10.1186/1471-2164-8-321
- Leekcharoenphon, P., Nielsen, E. M., Kaas, R. S., Lund, O., and Aarestrup, F. M. (2014). Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PLoS ONE* 9:e87991. doi: 10.1371/journal.pone.0087991
- Leopold, S. R., Magrini, V., Holt, N. J., Shaikh, N., Mardis, E. R., Cagno, J., et al. (2009). A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8713–8718. doi: 10.1073/pnas.0812949106
- Ma, B., Tromp, J., and Li, M. (2002). PatternHunter: faster and more sensitive homology search. *Bioinformatics* 18, 440–445. doi: 10.1093/bioinformatics/18.3.440
- Maddison, W. P., and Maddison, D. R. (2015). *Mesquite: a Modular System for Evolutionary Analysis*. Version 3.04.
- Makino, K., Ishii, K., Yasunaga, T., Hattori, M., Yokoyama, K., Yutsudo, C. H., et al. (1998). Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res.* 5, 1–9. doi: 10.1093/dnares/5.1.1
- Manning, S. D., Motiwala, A. S., Springman, A. C., Qi, W., Lacher, D. W., Ouellette, L. M., et al. (2008). Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4868–4873. doi: 10.1073/pnas.0710834105
- Manning, S. D., Whittam, T. S., and Steinsland, H. (2006). *EcMLST* [Online]. Available online at: <http://www.shigatox.net> (Accessed September 25 2014).
- Mellor, G. E., Fegan, N., Gobius, K. S., Smith, H. V., Jennison, A. V., D'astek, B. A., et al. (2015). Geographically distinct *Escherichia coli* O157 isolates differ by lineage, Shiga toxin genotype, and total shiga toxin production. *J. Clin. Microbiol.* 53, 579–586. doi: 10.1128/JCM.01532-14

- Meng, J., Zhao, S., Doyle, M. P., and Joseph, S. W. (1998). Antibiotic resistance of *Escherichia coli* O157:H7 and O157:NM isolated from animals, food, and humans. *J. Food Prot.* 61, 1511–1514.
- Meng, J., Zhao, S., Zhao, T., and Doyle, M. P. (1995). Molecular characterisation of *Escherichia coli* O157:H7 isolates by pulsed-field gel electrophoresis and plasmid DNA analysis. *J. Med. Microbiol.* 42, 258–263. doi: 10.1099/00222615-42-4-258
- Morelli, G., Song, Y., Mazzoni, C. J., Eppinger, M., Roumagnac, P., Wagner, D. M., et al. (2010). *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Publish. Group* 42, 1140–1143. doi: 10.1038/ng.705
- Munns, K. D., Zaheer, R., Xu, Y., Stanford, K., Laing, C. R., Gannon, V. P., et al. (2016). Comparative Genomic Analysis of *Escherichia coli* O157:H7 isolated from super-shedder and low-shedder cattle. *PLoS ONE* 11:e0151673. doi: 10.1371/journal.pone.0151673
- Myers, G. S., Mathews, S. A., Eppinger, M., Mitchell, C., O'Brien, K. K., White, O. R., et al. (2009). Evidence that human *Chlamydia pneumoniae* was zoonotically acquired. *J. Bacteriol.* 191, 7225–7233. doi: 10.1128/JB.00746-09
- Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.
- Ogura, Y., Mondal, S. I., Islam, M. R., Mako, T., Arisawa, K., Katsura, K., et al. (2015). The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Sci. Rep.* 5:16663. doi: 10.1038/srep16663
- Osterholm, M. T. (2015). Editorial commentary: the detection of and response to a foodborne disease outbreak: a cautionary tale. *Clin. Infect. Dis.* 61, 910–911. doi: 10.1093/cid/civ434
- Ostroff, S. M., Tarr, P. I., Neill, M. A., Lewis, J. H., Hargrett-Bean, N., and Kobayashi, J. M. (1989). Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. *J. Infect. Dis.* 160, 994–998. doi: 10.1093/infdis/160.6.994
- Perna, N. T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533. doi: 10.1038/35054089
- Persson, S., Olsen, K. E., Ethelberg, S., and Scheutz, F. (2007). Subtyping method for *Escherichia coli* shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J. Clin. Microbiol.* 45, 2020–2024. doi: 10.1128/JCM.02591-06
- Pettengill, J. B., Luo, Y., Davis, S., Chen, Y., Gonzalez-Escalona, N., Ottesen, A., et al. (2014). An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. *PeerJ* 2:e620. doi: 10.7717/peerj.620
- Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and its applications. *Genomics Proteomics Bioinformatics* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002
- Riordan, J. T., Viswanath, S. B., Manning, S. D., and Whittam, T. S. (2008). Genetic differentiation of *Escherichia coli* O157:H7 clades associated with human disease by real-time PCR. *J. Clin. Microbiol.* 46, 2070–2073. doi: 10.1128/JCM.00203-08
- Rusconi, B., and Eppinger, M. (2016). “Whole genome sequence typing for strategies for enterohemorrhagic *Escherichia coli* of the O157:H7 serotype,” in *The Handbook of Microbial Bioresources*, eds V. K. Gupta, G. D. Sharma, M. G. Tuohy, and R. Gaur (Wallingford, CT: Commonwealth Agricultural Bureaux International (CABI)), 656.
- Russo, L. M., Melton-Celsa, A. R., and O'Brien, A. D. (2016). Shiga Toxin (Stx) Type 1a reduces the Oral Toxicity of Stx Type 2a. *J. Infect. Dis.* 213, 1271–1279. doi: 10.1093/infdis/jiv557
- Sadiq, M. S., Hazen, T. H., Rasko, D. A., and Eppinger, M. (2014). “EHEC genomics: Past, Present, and Future,” in *Enterohemorrhagic *Escherichia coli* and other shiga Toxin-Producing *E. coli**, eds V. Sperandio and C. J. Hovde (Herndon, VA: ASM Press), 55–71.
- Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* 34, 374–378.
- Sahl, J. W., Caporaso, J. G., Rasko, D. A., and Keim, P. (2014). The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2:e332. doi: 10.7717/peerj.332
- Samadpour, M., Stewart, J., Steingart, K., Addy, C., Louderback, J., McGinn, M., et al. (2002). Laboratory investigation of an *E. coli* O157:H7 outbreak associated with swimming in Battle Ground Lake, Vancouver, Washington. *J. Environ. Health* 64, 16–20. 26, 25.
- Sanjar, F., Hazen, T. H., Shah, S. M., Koenig, S. S. K., Agrawal, S., Daugherty, S., et al. (2014). Genome sequence of *Escherichia coli* O157:H7 Strain 2886-75, associated with the first reported case of human infection in the united states. *Genome Announc.* 2:e01120-13. doi: 10.1128/genomeA.01120-13
- Sanjar, F., Rusconi, B., Hazen, T. H., Koenig, S. S., Mammel, M. K., Feng, P. C., et al. (2015). Characterization of the pathogenome and phylogenomic classification of enteropathogenic *Escherichia coli* of the O157:non-H7 serotypes. *Pathog. Dis.* 73:ftv033. doi: 10.1093/femspd/ftv033
- Schatz, M. C., and Phillippy, A. M. (2012). The rise of a digital immune system. *Gigascience* 1:4. doi: 10.1186/2047-217X-1-4
- Scheutz, F., Teel, L. D., Beutin, L., Piérard, D., Buvens, G., Karch, H., et al. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Serra-Moreno, R., Jofre, J., and Muniesa, M. (2008). The CI repressors of Shiga toxin-converting prophages are involved in coinfection of *Escherichia coli* strains, which causes a down regulation in the production of Shiga toxin 2. *J. Bacteriol.* 190, 4722–4735. doi: 10.1128/JB.00069-08
- Shaikh, N., and Tarr, P. I. (2003). *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J. Bacteriol.* 185, 3596–3605. doi: 10.1128/JB.185.12.3596-3605.2003
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Smith, D. L., Rooks, D. J., Fogg, P. C., Darby, A. C., Thomson, N. R., McCarthy, A. J., et al. (2012). Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics* 13:311. doi: 10.1186/1471-2164-13-311
- Strachan, N. J., Rotariu, O., Lopes, B., MacRae, M., Fairley, S., Laing, C., et al. (2015). Whole genome sequencing demonstrates that geographic variation of *Escherichia coli* O157 genotypes dominates host association. *Sci. Rep.* 5:14145. doi: 10.1038/srep14145
- Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039
- Tarr, P. I., Gordon, C. A., and Chandler, W. L. (2005). Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 365, 1073–1086. doi: 10.1016/s0140-6736(05)71144-2
- Tesh, V. L. (2010). Induction of apoptosis by Shiga toxins. *Future Microbiol.* 5, 431–453. doi: 10.2217/fmb.10.4
- Tesh, V. L., Burris, J. A., Owens, J. W., Gordon, V. M., Wadolkowski, E. A., O'Brien, A. D., et al. (1993). Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. *Infect. Immun.* 61, 3392–3402.
- Tilden, J. Jr., Young, W., Mcnamara, A. M., Custer, C., Boesel, B., Lambert-Fair, M. A., et al. (1996). A new route of transmission for *Escherichia coli*: infection from dry fermented salami. *Am. J. Public Health* 86, 1142–1145. doi: 10.2105/AJPH.86.8_Pt_1.1142
- Timme, R. E., Allard, M. W., Luo, Y., Strain, E., Pettengill, J., Wang, C., et al. (2012). Draft genome sequences of 21 *Salmonella enterica* serovar enteritidis strains. *J. Bacteriol.* 194, 5994–5995. doi: 10.1128/JB.01289-12
- Toro, M., Rump, L. V., Cao, G., Meng, J., Brown, E. W., and Gonzalez-Escalona, N. (2015). Simultaneous presence of insertion sequence-excision enhancer (IEE) and insertion sequence IS629 correlates with increased diversity and virulence in Shiga-toxin producing *Escherichia coli* (STEC). *J. Clin. Microbiol.* 53, 3466–3473. doi: 10.1128/JCM.01349-15
- Turabelidze, G., Lawrence, S. J., Gao, H., Sodergren, E., Weinstock, G. M., Abubucker, S., et al. (2013). Precise dissection of an *Escherichia coli* O157:H7 outbreak by single nucleotide polymorphism analysis. *J. Clin. Microbiol.* 51, 3950–3954. doi: 10.1128/JCM.01930-13
- Tuttle, J., Gomez, T., Doyle, M. P., Wells, J. G., Zhao, T., Tauxe, R. V., et al. (1999). Lessons from a large outbreak of *Escherichia coli* O157:H7 infections: insights into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiol. Infect.* 122, 185–192. doi: 10.1017/S0950268898001976

- Uhlich, G. A., Chen, C. Y., Cottrell, B. J., Hofmann, C. S., Dudley, E. G., Strobaugh, T. P. Jr., et al. (2013). Phage insertion in *mlrA* and variations in *rpoS* limit curli expression and biofilm formation in *Escherichia coli* serotype O157:H7. *Microbiology* 159, 1586–1596. doi: 10.1099/mic.0.066118-0
- Underwood, A. P., Dallman, T., Thomson, N. R., Williams, M., Harker, K., Perry, N., et al. (2013). Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. *J. Clin. Microbiol.* 51, 232–237. doi: 10.1128/JCM.01696-12
- Vogeleer, P., Tremblay, Y. D., Mafu, A. A., Jacques, M., and Harel, J. (2014). Life on the outside: role of biofilms in environmental persistence of Shiga-toxin producing *Escherichia coli*. *Front. Microbiol.* 5:317. doi: 10.3389/fmicb.2014.00317
- Vogler, A. J., Chan, F., Wagner, D. M., Roumagnac, P., Lee, J., Nera, R., et al. (2011). Phylogeography and molecular epidemiology of *Yersinia pestis* in Madagascar. *PLoS Negl. Trop. Dis.* 5:e1319. doi: 10.1371/journal.pntd.0001319
- Wang, J., Stephan, R., Power, K., Yan, Q., Hächler, H., and Fanning, S. (2014a). Nucleotide sequences of 16 transmissible plasmids identified in nine multidrug-resistant *Escherichia coli* isolates expressing an ESBL phenotype isolated from food-producing animals and healthy humans. *J. Antimicrob. Chemother.* 69, 2658–2668. doi: 10.1093/jac/dku206
- Wang, J., Stephan, R., Zurfluh, K., Hächler, H., and Fanning, S. (2014b). Characterization of the genetic environment of *bla* ESBL genes, integrons and toxin-antitoxin systems identified on large transferrable plasmids in multi-drug resistant *Escherichia coli*. *Front. Microbiol.* 5:716. doi: 10.3389/fmicb.2014.00716
- Wilgenbusch, J. C., and Swofford, D. (2003). Inferring evolutionary trees with PAUP*. *Curr. Protoc. Bioinformatics Chapter 6, Unit 6 4*. doi: 10.1002/0471250953.bi0604s00
- Wong, C. S., Mooney, J. C., Brandt, J. R., Staples, A. O., Jelacic, S., Boster, D. R., et al. (2012). Risk factors for the hemolytic uremic syndrome in children infected with *Escherichia coli* O157:H7: a multivariable analysis. *Clin. Infect. Dis.* 55, 33–41. doi: 10.1093/cid/cis299
- Wright, M. S., Perez, F., Brinkac, L., Jacobs, M. R., Kaye, K., Cober, E., et al. (2014). Population structure of KPC-producing *Klebsiella pneumoniae* isolates from midwestern U.S. hospitals. *Antimicrob. Agents Chemother.* 58, 4961–4965. doi: 10.1128/AAC.00125-14
- Xiong, Y., Wang, P., Lan, R., Ye, C., Wang, H., Ren, J., et al. (2012). A novel *Escherichia coli* O157:H7 clone causing a major hemolytic uremic syndrome outbreak in China. *PLoS ONE* 7:e36144. doi: 10.1371/journal.pone.0036144
- Yin, S., Rusconi, B., Sanjar, F., Goswami, K., Xiaoli, L. Z., Eppinger, M., et al. (2015). *Escherichia coli* O157:H7 strains harbor at least three distinct sequence types of Shiga toxin 2a-converting phages. *BMC Genomics* 16:733. doi: 10.1186/s12864-015-1934-1
- Zhang, W., Qi, W., Albert, T. J., Motiwala, A. S., Alland, D., Hyttia-Trees, E. K., et al. (2006). Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. *Genome Res.* 16, 757–767. doi: 10.1101/gr.4759706
- Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A. K., et al. (2007). Genome evolution in major *Escherichia coli* O157:H7 lineages. *BMC Genomics* 8:121. doi: 10.1186/1471-2164-8-121
- Zhao, S., White, D. G., Friedman, S. L., Glenn, A., Blickenstaff, K., Ayers, S. L., et al. (2008). Antimicrobial resistance in *Salmonella enterica* serovar Heidelberg isolates from retail meats, including poultry, from 2002 to 2006. *Appl. Environ. Microbiol.* 74, 6656–6662. doi: 10.1128/AEM.01249-08
- Zhou, S., Bechner, M. C., Place, M., Churas, C. P., Pape, L., Leong, S. A., et al. (2007). Validation of rice genome sequence by optical mapping. *BMC Genomics* 8:278. doi: 10.1186/1471-2164-8-278
- 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
- Zhou, Z., Li, X., Liu, B., Beutin, L., Xu, J., Ren, Y., et al. (2010). Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS ONE* 5:e8700. doi: 10.1371/journal.pone.0008700

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

Copyright © 2016 Rusconi, Sanjar, Koenig, Mammel, Tarr and Eppinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

