

RESEARCH ARTICLE

# Comparative genomics reveals differences in mobile virulence genes of *Escherichia coli* O103 pathotypes of bovine fecal origin

Lance W. Noll<sup>1</sup>, Jay N. Worley<sup>2</sup>, Xun Yang<sup>2</sup>, Pragathi B. Shridhar<sup>1</sup>, Justin B. Ludwig<sup>1</sup>, Xiaorong Shi<sup>1</sup>, Jianfa Bai<sup>3</sup>, Doina Caragea<sup>4</sup>, Jianghong Meng<sup>2</sup>, T. G. Nagaraja<sup>1\*</sup>

**1** Department of Diagnostic Medicine/Pathobiology, Kansas State University, Manhattan, Kansas, United States of America, **2** Joint Institute for Food Safety and Applied Nutrition and Department of Nutrition and Food Science, University of Maryland, College Park, Maryland, United States of America, **3** Veterinary Diagnostic Laboratory, Kansas State University, Manhattan, Kansas, United States of America, **4** Department of Computing and Information Sciences, Kansas State University, Manhattan, Kansas, United States of America

\* [tnagaraj@vet.k-state.edu](mailto:tnagaraj@vet.k-state.edu)



**OPEN ACCESS**

**Citation:** Noll LW, Worley JN, Yang X, Shridhar PB, Ludwig JB, Shi X, et al. (2018) Comparative genomics reveals differences in mobile virulence genes of *Escherichia coli* O103 pathotypes of bovine fecal origin. PLoS ONE 13(2): e0191362. <https://doi.org/10.1371/journal.pone.0191362>

**Editor:** Chitrita DebRoy, The Pennsylvania State University, UNITED STATES

**Received:** September 5, 2017

**Accepted:** January 3, 2018

**Published:** February 1, 2018

**Copyright:** © 2018 Noll et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was supported in part by the Agriculture and Food Research Initiative Competitive Grant no. 2012-68003-30155 from the United States Department of Agriculture National Institute of Food and Agriculture (to TGN). The funders had no role in the study design, data collection and analyses, preparation of manuscript or decision to publish.

## Abstract

*Escherichia coli* O103, harbored in the hindgut and shed in the feces of cattle, can be enterohemorrhagic (EHEC), enteropathogenic (EPEC), or putative non-pathotype. The genetic diversity particularly that of virulence gene profiles within O103 serogroup is likely to be broad, considering the wide range in severity of illness. However, virulence descriptions of the *E. coli* O103 strains isolated from cattle feces have been primarily limited to major genes, such as Shiga toxin and intimin genes. Less is known about the frequency at which other virulence genes exist or about genes associated with the mobile genetic elements of *E. coli* O103 pathotypes. Our objective was to utilize whole genome sequencing (WGS) to identify and compare major and putative virulence genes of EHEC O103 (positive for Shiga toxin gene, *stx1*, and intimin gene, *eae*; n = 43), EPEC O103 (negative for *stx1* and positive for *eae*; n = 13) and putative non-pathotype O103 strains (negative for *stx* and *eae*; n = 13) isolated from cattle feces. Six strains of EHEC O103 from human clinical cases were also included. All bovine EHEC strains (43/43) and a majority of EPEC (12/13) and putative non-pathotype strains (12/13) were O103:H2 serotype. Both bovine and human EHEC strains had significantly larger average genome sizes ( $P < 0.0001$ ) and were positive for a higher number of adherence and toxin-based virulence genes and genes on mobile elements (prophages, transposable elements, and plasmids) than EPEC or putative non-pathotype strains. The genome size of the three pathotypes positively correlated ( $R^2 = 0.7$ ) with the number of genes carried on mobile genetic elements. Bovine strains clustered phylogenetically by pathotypes, which differed in several key virulence genes. The diversity of *E. coli* O103 pathotypes shed in cattle feces is likely reflective of the acquisition or loss of virulence genes carried on mobile genetic elements.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Enterohemorrhagic *Escherichia coli* (EHEC) carry one or both phage-encoded Shiga toxin genes (*stx1* and *stx2*) and the attaching and effacing gene (*eae*), which is harbored in the chromosomal-encoded locus of enterocyte effacement (LEE) pathogenicity island. Among EHEC pathotypes, O157:H7 serotype is most frequently associated with human foodborne illness. However, Centers for Disease Control and Prevention (CDC) rank O103 as the second most common serogroup, next to O26, identified in laboratory confirmed non-O157 EHEC infections in the U.S. [1]. In human EHEC infections, disease outcomes can range from mild to bloody diarrhea (hemorrhagic colitis) to more serious complications, such as hemolytic uremic syndrome (HUS), and even death [2]. Differences in disease-causing potential, particularly the ability to cause serious complications, are attributed to differences in virulence of EHEC strains [3]. In addition to the major virulence factors, which include Shiga toxins and LEE gene-encoded proteins, other virulence attributes, including known putative virulence factors, contribute to the development, progression, and outcome of the disease [4–6]. Enteropathogenic *E. coli* (EPEC), including EPEC O103, do not carry *stx* genes; however, they possess *eae* and other virulence genes to cause attaching and effacing lesions that can result in mild to severe diarrhea, or even death, particularly in children [7, 8]. Strains within the EPEC pathotype are further characterized as typical or atypical, depending on presence or absence, respectively, of the EPEC adherence factor (EAF) plasmid [9]. The loss of the *stx* gene(s), a frequently reported event [10, 11], can transform an EHEC into an EPEC pathotype. These major pathotype-defining mobile virulence genes have been well studied, but less is known about how other mobile elements contribute to the overall virulence diversity in O103 serogroup. Some strains of *E. coli* O103 carry neither Shiga toxin nor intimin genes, possibly a non-pathotype; even less is known about the virulence profiles of these strains. Cattle have been shown to harbor EHEC, EPEC and putative non-pathotype O103 in the hindgut and shed them in the feces [12]. We hypothesize that the diversity of O103 pathotypes harbored and shed in the feces of cattle is reflective of the loss or acquisition of genes carried on mobile genetic elements.

Whole genome sequencing (WGS) has been used to characterize the virulence gene profiles of EHEC O157 [13], identify phylogenetic relationships between EHEC O157 and non-O157 serotypes [14–18] as well as discover novel virulence determinants [19]. However, differences in virulence gene profiles and phylogenetic relationships of O103 pathotypes of bovine origin are less characterized [20]. Therefore, our objectives were to utilize WGS to identify and compare major and putative virulence genes, particularly genes located on mobile elements, of bovine and human clinical EHEC O103, bovine EPEC O103, and putative non-pathotype O103 strains and analyze phylogenetic relationships among the strains.

## Materials and methods

### Strains

The Institutional Animal Care and Use Committee at Kansas State University approved the research that resulted in the strains that were used in the study. The bovine EHEC strains investigated in this study were isolated from cattle feces from several feedlots in the Midwest region of the US [12, 21, 22]. Sixty-nine bovine O103 strains, previously identified by endpoint PCR [23] as positive for *stx1* (Shiga toxin 1) and *eae* (intimin) (bovine EHEC; n = 43), negative for *stx1* and positive for *eae* (bovine EPEC; n = 13) and negative for both *stx1* and *eae* (bovine putative non-pathotype; n = 13) were used in the study. Human clinical O103 strains positive for *stx1* and *eae* (human EHEC; n = 6) were included in the study for comparison. The strains were cultured onto Tryptone soy agar (TSA; BD Difco, Sparks, MD) slants and

shipped overnight in cold storage to the University of Maryland for whole genome sequencing.

### DNA preparation and whole genome sequencing

The O103 strains from TSA slants were streaked onto blood agar (Remel, Lenexa, KS) and then subcultured in tryptone soy broth (BD Difco, Sparks, MD). Bacterial DNA from overnight culture was extracted from each strain using the DNeasy Blood and Tissue Kit with the QIAcube robotic workstation (Qiagen, Germantown, MD). The genomes were sequenced using an Illumina MiSeq platform (Illumina, San Diego, CA) at approximately 37x average coverage. Genomic libraries were constructed using Nextera XT DNA Library Preparation Kit and MiSeq Reagent Kits v2 (500 Cycles) (Illumina, Inc.). *De novo* genome assembly was performed using SPAdes 3.6.0 [24].

### Genomic analysis

Draft genomes were annotated using Rapid Annotation using Subsystem Technology (RAST version 2.0 - <http://rast.nmpdr.org/>; [25]), a web-based service commonly used for annotation of draft bacterial genomes [26, 27]. RAST applies the Fellowship for Interpretation of Genomes (FIG) subsystem approach to rapidly call and annotate genes, then uses high-throughput comparative analysis and a collection of expertly curated databases to categorize genes, based on the functional role they perform, into subsystems. Average number of genes located on mobile elements (prophages, transposable elements and plasmids), and genes related to virulence, disease and defense were determined, using RAST, for each of the O103 subgroups (bovine EHEC, human EHEC, bovine EPEC and bovine putative non-pathotype). Genomic sequencing data in this study exceeded the minimum criteria for analysis that RAST requires of each genome: at least 10x coverage (using 454 pyrosequencing) and 70% of assembled sequences in contigs > 20,000 base pairs. Serotype identity, virulence and plasmid make-up of the 75 strains were determined using default parameters of Center for Genomic Epidemiology Serotype-Finder 1.1 (<https://cge.cbs.dtu.dk/services/SerotypeFinder/>) [28], Virulence Finder 1.4 (<https://cge.cbs.dtu.dk/services/VirulenceFinder/>) [29], and PlasmidFinder 1.3 [30] programs, respectively. Prophage sequences of the 75 strains were determined using Phage Search Tool Enhanced Release (PHASTER; <http://phaster.ca/>) [31, 32]; intact, and questionable prophage sequences, defined by PHASTER scores of >90 and 70–90, respectively, were included in analysis. The complete genome of EHEC O103:H2 strain 12009 (GenBank accession no. AP010958.1; <https://www.ncbi.nlm.nih.gov/nuccore/AP010958.1>) and 12009 plasmid pO103 DNA (GenBank accession no. NC\_013354.1; [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_013354.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_013354.1)), a classical O103 reference strain of clinical origin used in many O103 genomic studies [14, 33, 34], was tested with Virulence Finder 1.4, ResFinder 2.1, Plasmid Finder 1.3 and PHASTER as a control for comparison. The complete genomes EHEC O157:H7 Sakai (GenBank accession no. BA000007.2; <https://www.ncbi.nlm.nih.gov/nuccore/BA000007.2>) and EHEC O157:H7 EDL933 (GenBank accession no. CP008957.1; <https://www.ncbi.nlm.nih.gov/nuccore/CP008957.1>) and their associated plasmids (Sakai plasmid pO157: GenBank accession no. NC\_002128.1, [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_002128.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_002128.1); Sakai plasmid pOSAK1: GenBank accession no. NC\_002127.1, [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_002127.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_002127.1); EDL933 plasmid pO157: GenBank accession no. AF074613.1, <https://www.ncbi.nlm.nih.gov/nuccore/AF074613.1>) were also tested for comparison. Parsnp v1.2 (<http://harvest.readthedocs.io/en/latest/content/parsnp.html>) [35] was used for core genome alignment of the 75 strains and subsequent construction of a maximum likelihood tree. For improved visualization, a proportional branch transformation of the output file (.tree) from Parsnp was performed using FigTree 1.4 software (<http://tree.bio>.

[ed.ac.uk/software/figtree/](http://ed.ac.uk/software/figtree/)) [36] and bootstrap values were reported for each branch. Representative strains, based on clustering patterns observed in the phylogenetic tree, were chosen as input for BLAST Ring Image Generator software (BRIG v0.95 - <https://sourceforge.net/projects/brig/>) [37]. The BRIG plot displays similarities and differences between the draft genome nucleotide sequence identities of target strains, represented by concentric rings, to the genome identity of a chosen reference strain, identified in the center of the BRIG plot. The complete genome of EHEC O103:H2 strain 12009 was used as a BRIG plot reference. The nucleotide sequence (45,325 bp) of the LEE pathogenicity island (GenBank accession no.: AF071034.1; <https://www.ncbi.nlm.nih.gov/nuccore/AF071034.1>) of human clinical EHEC O157:H7 EDL933 strain [37] was mapped to the BRIG plot for comparison of LEE between the target strains.

### Statistical analysis

A single factor analysis of variance (ANOVA) test was performed to determine whether average genome size, and average number of extra-chromosomal genes and virulence, disease and defense genes were significantly different among the four subgroups (bovine EHEC, human EHEC, EPEC and putative non-pathotype). If means were significantly different ( $P \leq 0.01$ ), Tukey adjustment for multiple comparisons was performed, using SAS 9.4 with Proc Glimmix, to test each pairwise comparison for significant differences ( $P \leq 0.01$ ).

### Nucleotide sequence accession numbers

Draft genome sequences of the 75 *E. coli* O103 strains are available in GenBank and their accession numbers are listed in Tables in [S1](#), [S2](#) and [S3](#) Tables.

## Results

Sixty-nine bovine O103 strains, that belonged to three subgroups, EHEC ( $n = 43$ ), EPEC ( $n = 13$ ) and putative non-pathotype ( $n = 13$ ) and six human clinical EHEC O103 strains were included in the study. All bovine EHEC strains (43/43; 100%) and a majority of EPEC (12/13; 92.3%) and putative non-pathotype strains (12/13; 92.3%) were O103:H2 serotype. The two remaining strains of EPEC (1/13) and putative non-pathotype (1/13) were O103:H11 and O103:H16 serotypes, respectively. Four of the six human EHEC strains were O103:H11 and two were O103:H2 serotype.

### RAST subsystem summary

Genome size range of bovine (5.32–5.79 Mb) and human EHEC (5.43–5.77 Mb) subgroups were similar ([Table 1](#)). However, both bovine and human EHEC subgroups had significantly larger average genome sizes ( $P \leq 0.0001$ ) compared to EPEC or putative non-pathotype subgroups. Average genome size was similar between EPEC and putative non-pathotype subgroups. However, one of the bovine EPEC O103:H11 strains (2013-3-492A) had a similar genome size (5.67 Mb) to that of other EHEC strains.

Overall, the number of genes in the category of virulence, disease and defense was comparable for all 75 strains tested ([Table 1](#)), with no significant differences observed in the mean number of genes among the O103 subgroups. However, the number of genes on mobile elements (prophages, transposable elements, and plasmids) varied considerably among O103 subgroups and among serotypes within subgroups. Strains belonging to bovine and human EHEC subgroups had a significantly higher ( $P \leq 0.001$ ) number of mobile genes compared to EPEC and putative non-pathotype subgroups. Average number of mobile genes was not significantly different between bovine and human EHEC subgroups or between EPEC and putative



**Table 1. Average genome size, guanine-cytosine (GC) content, and number of contigs and average number of extra-chromosomal genes, virulence, disease and defense genes and plasmids of enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/ea*e negative) *Escherichia coli* O103 strains of bovine and human origin.**

Genome size and gene categories <sup>†</sup>	Host origin, pathotype and serotype (no. of strains tested)					
	Bovine EHEC	Human EHEC		Bovine EPEC	Bovine putative non-pathotype	
	O103:H2 (n = 43)	O103:H2 (n = 2)	O103:H11 (n = 4)	O103:H2 (n = 12)	O103:H11 (n = 1)	O103:H2 (n = 12)
Genome size (Mb)	5.47 (5.32–5.79)	5.45 (5.43–5.46)	5.61 (5.52–5.77)	5.22 (5.16–5.33)	5.67	5.26 (5.21–5.33)
GC content (%)	50.58 (50.5–50.6)	50.6 (50.6–50.6)	50.43 (50.4–50.5)	50.54 (50.5–50.6)	50.5	50.44 (50.4–50.5)
Contigs	312 (181–398)	350 (339–360)	432 (412–453)	204 (137–268)	406	140 (99–172)
Virulence, disease and defense	113 (111–124)	111 (111–111)	112 (109–121)	114 (113–117)	109	115 (115–116)
Prophages, transposable elements and plasmids	260 (221–351)	265 (260–270)	273 (256–292)	157(137–213)	289	128 (100–157)
Plasmids	2.9 (1–6)	2.5 (2–3)	4 (3–6)	2 (0–4)	5	3 (3–4)

<sup>†</sup>Genome sizes, GC content, contigs, virulence, disease and defense and mobile element (prophages, transposable elements and plasmids) data were determined using Rapid Annotation Using Subsystem Technology (RAST; [25]). Plasmid data was determined using PlasmidFinder 1.3 [30].

<https://doi.org/10.1371/journal.pone.0191362.t001>

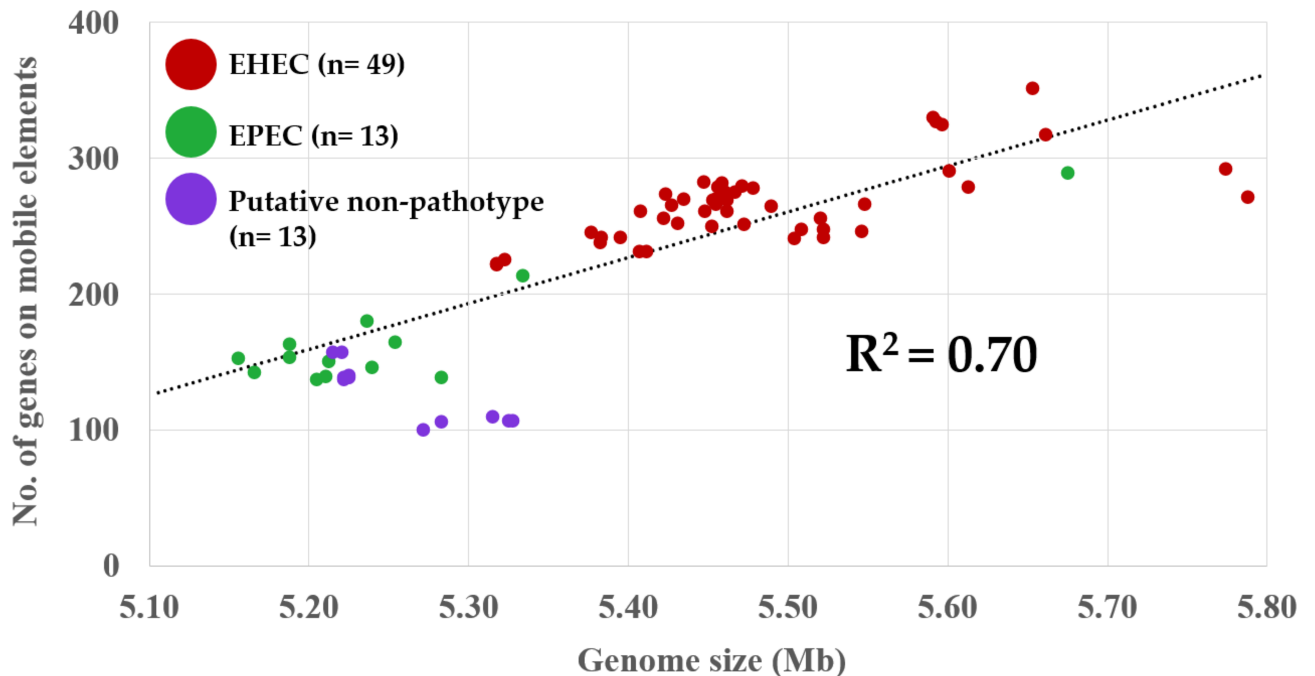
non-pathotype subgroups. The bovine EHEC strains possessed the widest range in the number of genes on mobile elements (221–351). Similarly, wide ranges were observed in bovine EPEC strains (137–289 genes) and bovine putative non-pathotype strains (100–157 genes), but not in human EHEC strains (256–292 genes). Mobile gene counts above 300 were only observed in a few bovine EHEC strains (4/43), and one bovine EHEC strain (2014-5-933A) had 351 mobile genes, nearly 60 more than the highest number in strains of the human EHEC subgroup. Furthermore, the one bovine EPEC O103:H11 (strain 2013-3-492A) that had a similar genome size as EHEC pathotype had 289 mobile genes; 76 more mobile genes than the highest number in strains within the EPEC O103:H2 subgroup.

A strong correlation ( $R^2 = 0.70$ ) was observed between genome size vs. number of genes on mobile elements for the 75 strains (Fig 1). The EHEC strains had larger genome size and higher number of genes on mobile elements compared to EPEC and putative non-pathotype strains. The EPEC O103:H11 strain (2013-3-492A) appeared to be an EPEC outlier, with genome size and number of genes on mobile elements closer to those of the EHEC O103 strains (Fig 1).

### Virulence genes

Virulence genes with >90% sequence homology were considered positive in a genome. The complete virulence gene profiles of each genome are shown in tables in S1, S2 and S3 Tables. All EHEC strains were positive for Shiga toxin 1a (*stx1a*) subtype. On average, bovine and human EHEC strains were positive for more virulence genes than EPEC strains; putative non-pathotype strains were negative for all LEE encoded, non-LEE encoded, and pO157 plasmid-encoded genes (Table 2).

Among LEE-encoded genes, all EHEC and EPEC strains were positive for *ea*e, translocated intimin receptor protein (*tir*), and type III secretion effectors (*espA* and *espB*), but a small number of bovine EHEC (4/43) and EPEC O103:H2 strains (3/12) were negative for type III secretion effector gene, *espF* (Table 2). All EHEC/EPEC O103:H2 and O103:H11 serotypes were positive for *ea*e-epsilon and *ea*e-beta1 subtypes, respectively. Other phage-encoded type III secretion effector genes (*cif*, *espJ*, and *tccP*) were present in all human EHEC O103:H2 strains but were present at varying proportions for other EHEC and EPEC O103 subgroups. Non-LEE encoded effectors A (*nleA*) and B (*nleB*) were present in all EHEC strains, in the EPEC O103:H11 strain, but also in a majority of EPEC O103:H2 strains (6/12 for *nleA* and 10/



**Fig 1.** Scatterplot of genome sizes and number of genes on mobile elements<sup>†</sup> of 75 strains of enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/ea*e negative) *Escherichia coli* O103. <sup>†</sup>Genome sizes and number of genes located on mobile elements (prophages, transposable elements and plasmids) were determined using Rapid Annotation Using Subsystem Technology (RAST; [25]).

<https://doi.org/10.1371/journal.pone.0191362.g001>

12 for *nleB*). The *nleC* gene, absent in two human EHEC O103:H2 strains, was present in all human EHEC O103:H11 strains (4/4) and also in over half of bovine EHEC O103:H2 strains (23/43; 53.5%).

Among pO157 plasmid-encoded genes (*ehxA*, *espP*, *etpD*, *katP* and *toxB*), enterohemolysin (*ehxA*) and extracellular serine protease (*espP*) were present in most, but not all EHEC and EPEC strains (Table 2). Conversely, toxin B gene (*toxB*), a homolog of EHEC factor for adherence gene (*efa1*), was found in only 2/6 (33.3%) human clinical EHEC and in only one bovine EHEC strain (2014-5-941B). The *efa1* gene, not encoded on the pO157 plasmid, was present in a higher proportion of EHEC strains (41/49; 83.7%), compared to *toxB*; interestingly, bovine EPEC O103:H11 strain was also positive for *efa1* gene (Table 3). All EPEC strains in this study were negative for the EAF plasmid.

The putative virulence genes that were present in the O103 strains are shown in Table 3. Of all adherence-based genes in EHEC and EPEC strains (Tables 2 and 3), only long polar fimbriae gene (*lpfA*) was present in putative non-pathotype strains. The *lpfA* gene was also present in all human EHEC O103:H11 strains (n = 4) and in the EPEC O103:H11 strain, but was not detected in O103:H2 strains within bovine and human EHEC and bovine EPEC subgroups or within any of the human EHEC control strains (O103:H2 12009, O157:H7 Sakai, O157:H7 EDL933). ABC transporter protein MchF (*mcfF*), MchC protein (*mchC*), Microcin H47 part of colicin H (*mchB*) and Microcin M part of colicin H (*mcmA*) genes were present in 5/12 (41.7%) bovine putative non-pathotype O103:H2 strains but absent in all other strains. The colicin M gene (*cma*) was found in 5 of 12 putative non-pathotype O103:H2 strains, but also in one bovine EHEC O103:H2 (strain 2014-5-1565C). Glutamic acid decarboxylase (*gad*) was present in all 75 strains. EAST-1 toxin gene (*astA*), encoding for an enterotoxin, was in all O103:H11 strains (human EHEC and bovine EPEC) in the study, and in a majority of bovine

**Table 2. Major chromosomal-, phage-, and plasmid-encoded virulence genes in enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/ae* negative) *Escherichia coli* O103 strains of bovine and human origin<sup>†</sup>.**

Origin	Protein and gene		Host origin, pathotype and serotype (no. of strains)						
			Bovine EHEC	Human EHEC		Bovine EPEC		Bovine putative non-pathotype	
			O103:H2	O103:H2	O103:H11	O103:H2	O103:H11	O103:H2	O103:H16
			(n = 43)	(n = 2)	(n = 4)	(n = 12)	(n = 1)	(n = 12)	(n = 1)
Locus of enterocyte effacement (LEE) encoded	Intimin	<i>eae</i>	43	2	4	12	1	0	0
	Translocated intimin receptor	<i>tir</i>	43	2	4	12	1	0	0
	Type III secretions system	<i>espA</i>	43	2	4	12	1	0	0
	Secreted protein B	<i>espB</i>	43	2	4	12	1	0	0
	Type III secretion system	<i>espF</i>	39	2	4	9	1	0	0
Non-LEE encoded	Effector A	<i>nleA</i>	43	2	4	6	1	0	0
	Effector B	<i>nleB</i>	43	2	4	10	1	0	0
	Effector C	<i>nleC</i>	23	0	4	0	0	0	0
Phage-encoded Shiga toxin	O157 FLY16, variant a	<i>stx1</i>	42	2	4	0	0	0	0
	<i>Shigella dysenteriae</i> 3818T	<i>stx1</i>	1	0	0	0	0	0	0
	Shiga toxin 1, subunit A, variant a	<i>stx1A</i>	43	2	4	0	0	0	0
	Shiga toxin 1, subunit B, variant a	<i>stx1B</i>	43	2	4	0	0	0	0
Phage-encoded type III secretion effectors	Cycle inhibiting factor	<i>cif</i>	43	2	4	0	1	0	0
	Type III secretion system effector	<i>espJ</i>	38	2	4	0	1	0	0
	Tir-cytoskeleton coupling protein	<i>tccP</i>	41	2	3	12	0	0	0
pO157 plasmid-encoded	Enterohaemolysin	<i>ehxA</i>	43	2	4	11	1	0	0
	Extracellular serine protease	<i>espP</i>	37	2	4	11	1	0	0
	Type II secretion protein	<i>etpD</i>	9	0	0	0	0	0	0
	Catalase peroxidase	<i>katP</i>	29	2	4	0	1	0	0
	Toxin B	<i>toxB</i>	1	1	1	0	1	0	0

<sup>†</sup>Virulence genes were determined using Virulence Finder 1.4 [29].

<https://doi.org/10.1371/journal.pone.0191362.t002>

EPEC O103:H2 strains (9/12), but not in any of the EHEC O103:H2 strains. Endonuclease colicin E2 gene (*celb*) was present in nearly half (20/43) of all bovine EHEC strains, and in the bovine EPEC O103:H11 strain, but absent from all other subgroups.

### Plasmid and prophage sequences

The complete plasmid replicon profiles of each genome are shown in tables S4, S5 and S6 Tables. Plasmid profiles exhibited some commonality among strains within an O103 subgroup but varied dramatically between subgroups. Plasmids from four incompatibility groups, including IncFIA(H11), IncFII(pRSB107), IncFII(pSE11), IncX1 and IncY replicons were present at varying proportions in bovine EPEC O103:H12 strains, but absent from all other subgroups (Table 4). Similarly, strains from bovine EHEC were positive for IncA/C2, IncFII (pCoo) (enterotoxigenic *E. coli* associated plasmid), IncI2 and IncN plasmid replicons, while other subgroups were negative for these plasmids sequences. A high proportion of bovine EHEC (19/43; 44.2%) and the bovine EPEC O103:H11 strains were positive for Col156 plasmid sequence, while strains from all remaining subgroups were negative for this plasmid

**Table 3. Putative virulence genes in enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/eae* negative) *Escherichia coli* O103 strains of bovine and human origin<sup>†</sup>.**

Protein and gene		Host origin, pathotype and serotype (no. of strains)						
		Bovine EHEC	Human EHEC		Bovine EPEC		Bovine putative non-pathotype	
		O103:H2	O103:H2	O103:H11	O103:H2	O103:H11	O103:H2	O103:H16
		(n = 43)	(n = 2)	(n = 4)	(n = 12)	(n = 1)	(n = 12)	(n = 1)
EHEC factor for adherence	<i>efa1</i>	35	2	4	0	1	0	0
Adherence protein	<i>iha</i>	19	1	4	2	1	0	0
EAST-1 heat-stable toxin	<i>astA</i>	0	0	4	9	1	0	0
SPATE	<i>espI</i>	23	0	0	3	0	0	0
Glutamic acid decarboxylase	<i>gad</i>	43	2	4	12	1	12	1
Increased serum survival	<i>Iss</i>	43	2	4	9	1	12	1
Long polar fimbriae	<i>lpfA</i>	0	0	4	0	1	12	1
Endonuclease colicin E2	<i>celb</i>	20	0	0	0	1	0	0
Colicin B	<i>cba</i>	29	0	4	0	1	5	0
Colicin M	<i>cma</i>	1	0	0	0	0	5	0
ABC transporter protein MchF	<i>mchF</i>	0	0	0	0	0	5	0
MchC protein	<i>mchC</i>	0	0	0	0	0	5	0
Microcin H47 part of colicin H	<i>mchB</i>	0	0	0	0	0	5	0
Microcin M part of colicin H	<i>mcmA</i>	0	0	0	0	0	5	0

<sup>†</sup>Virulence genes were determined using Virulence Finder 1.4 [29].

<https://doi.org/10.1371/journal.pone.0191362.t003>

sequence. Among the nineteen total plasmid types identified in the strains, nearly half (9/19; 47.4%) belonged to the IncF incompatibility family. The IncFIB (*E. coli* K-12) plasmid sequence was most prevalent among the 75 strains, found in 39/43 (90.7%) bovine EHEC strains and in all human EHEC (6/6) and O103:H2 putative non-pathotype strains (12/12). The IncFIB plasmid sequence was present in the bovine EPEC O103:H11 strain, but absent from all EPEC O103:H2 strains.

The complete prophage profiles of each genome are shown in tables S7, S8 and S9 Tables. The 75 strains were positive for 20 different prophages (Table 5). Bovine EHEC strains were positive for the most number of these prophages (15/20), followed by bovine EPEC (11/20) and human EHEC strains (8/20). Bovine putative non-pathotype strains were positive for the fewest number of these prophages (5/20). A high proportion of bovine EHEC (28/43; 65.1%), human EHEC (5/6; 83.3%), and bovine EPEC (5/13; 38.5%) were positive for *Enterobacteria* phage P88, while only 7.7% (1/13) of bovine putative non-pathotype strains were positive for this prophage (Table 5). Interestingly, 61.5% of bovine putative non-pathotype strains (8/13) and 62.8% of bovine EHEC strains were positive for *Shigella* phage SfII, compared to none of the bovine EPEC strains and only 2 of 6 human EHEC strains.

### Phylogenetic relationships

A maximum likelihood phylogenetic tree, based on core genome alignment of all 75 strains, was constructed using Parsnp v.1.2. The output file was proportional branch transformed using FigTree 1.4 (Fig 2). Overall, strains clustered according to pathotypes, with one notable exception: bovine EPEC O103:H11 strain (2013-3-492A) was more closely related to a human EHEC O103:H11 (strain KSU-74) than to any of the other bovine EPEC strains included in the study (Fig 2). All EPEC O103:H2 strains clustered together and putative non-pathotype strains exhibited a similar clustering. One human EHEC O103:H2 strain (KSU-72) was more

**Table 4. Number of enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/ea*e negative) *Escherichia coli* O103 strains of bovine and human origin positive for plasmids<sup>†</sup>.**

Plasmid replicon	Host origin, pathotype and serotype (no. isolates tested)						
	Bovine EHEC	Human EHEC		Bovine EPEC		Bovine putative non-pathotype	
	O103:H2 (n = 43)	O103:H2 (n = 2)	O103:H11 (n = 4)	O103:H2 (n = 12)	O103:H11 (n = 1)	O103:H2 (n = 12)	O103:H16 (n = 1)
Col156	19	0	0	0	1	0	0
ColRNAI	36	0	4	0	1	12	1
IncA/C2	3	0	0	0	0	0	0
IncB/O/K/Z	20	2	4	0	1	0	0
IncFIA(HI1)	0	0	0	2	0	0	0
IncFIB(AP001918)	39	2	4	0	1	12	0
IncFIC(FII)	0	0	0	0	0	12	0
IncFII	1	0	1	0	0	0	0
IncFII(29)	0	1	0	0	0	0	0
IncFII(pCoo)	3	0	0	0	0	0	0
IncFII(pHN7A8)	0	0	0	2	0	5	0
IncFII(pRSB107)	0	0	0	2	0	0	0
IncFII(pSE11)	0	0	0	9	0	0	0
IncI2	1	0	0	0	0	0	0
IncN	2	0	0	0	0	0	0
IncP	0	0	1	0	0	0	0
IncX1	0	0	0	2	0	0	0
IncY	0	0	0	7	0	0	0
p0111	2	0	2	0	1	0	0

<sup>†</sup>Plasmids were determined from whole genome sequences of strains using Plasmid Finder 1.3 [30].

<https://doi.org/10.1371/journal.pone.0191362.t004>

closely related to two bovine EHEC O103:H2 strains (2014-5-330A and 2014-5-332A) than to the other human EHEC O103:H2 strain (KSU-71) included in the study.

Based on clustering patterns in Fig 2, representative strains were selected from observed serotypes (O103:H2, O103:H11, and O103:H16) within each O103 subgroup (bovine EHEC, human EHEC, bovine EPEC, and bovine putative non-pathotype) as input for BLAST Ring Image Generator (BRIG) v0.95 [37]. The draft genomes of these target strains are represented by the concentric rings in the BRIG plot; any missing portions of these rings represent nucleotide sequences missing from the target strains in comparison to a central reference strain (EHEC O103:H2 strain 12009; Fig 3). Putative non-pathotype strains (2013-3-308C and 2013-3-111C) displayed the largest degree of sequence divergence to the reference strain. As expected, the LEE island (45,325 bp), which encodes for the *ea*e gene and other Type III secretion effectors, was present in all EHEC and EPEC strains, but absent in the putative non-pathotype strains. Interestingly, a relatively large unknown sequence (~40,000 bp) from the reference strain was present in 2/5 bovine EHEC O103:H2 strains (2013-3-174C, 2014-5-1565C) and in 1/3 human EHEC strains (KSU-72), but absent in all other EHEC, EPEC, and putative non-pathotype strains. It is worth noting that the three strains positive for the unknown sequence were not positive for any virulence genes not found in the remaining strains tested. Strains 2013-3-174C and 2014-5-1565C of bovine EHEC O103:H2 had higher sequence similarity with the human clinical O103:H2 reference strain than to any of the human clinical EHEC target strains.



**Table 5. Number of enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/ea*e negative) *Escherichia coli* O103 strains of bovine and human origin positive for prophages<sup>†</sup>.**

Prophage	Host origin, pathotype and serotype (no. isolates tested)						
	Bovine EHEC	Human EHEC		Bovine EPEC		Bovine putative non-pathotype	
	O103:H2	O103:H2	O103:H11	O103:H2	O103:H11	O103:H2	O103:H16
	(n = 43)	(n = 2)	(n = 4)	(n = 12)	(n = 1)	(n = 12)	(n = 1)
<i>Aeromonas</i> phage phiO18P	21	1	0	0	0	0	0
<i>Burkholderia</i> phage phiE255	0	0	0	0	0	5	0
<i>Enterobacteria</i> phage 933W	7	0	0	6	0	0	0
<i>Enterobacteria</i> phage BP-4795	8	0	0	2	0	0	0
<i>Enterobacteria</i> phage mEp043 c-1	8	1	0	0	0	0	0
<i>Enterobacteria</i> phage HK630	3	0	0	0	0	0	0
<i>Enterobacteria</i> phage lambda	19	2	0	4	0	0	0
<i>Enterobacteria</i> phage mEp460	3	0	0	8	0	0	0
<i>Enterobacteria</i> phage Mu	2	0	2	0	1	0	0
<i>Enterobacteria</i> phage P1	2	0	2	3	1	0	0
<i>Enterobacteria</i> phage P2	0	0	1	0	0	0	0
<i>Enterobacteria</i> phage P88	28	2	3	5	0	0	1
<i>Enterobacteria</i> phage Sfl01	0	0	0	2	0	0	0
<i>Escherichia</i> phage D108	2	0	0	0	0	0	0
<i>Escherichia</i> phage TL-2011b	8	0	0	8	0	0	0
<i>Enterobacteria</i> phage Fels-2	0	0	0	0	0	7	0
<i>Salmonella</i> phage SEN34	0	0	0	0	0	3	0
<i>Shigella</i> phage SflI	27	2	0	0	0	7	1
<i>stx2</i> -converting phage 1717	5	0	0	1	0	0	0
<i>stx2</i> converting phage vB_EcoP_24B	1	0	0	2	0	0	0

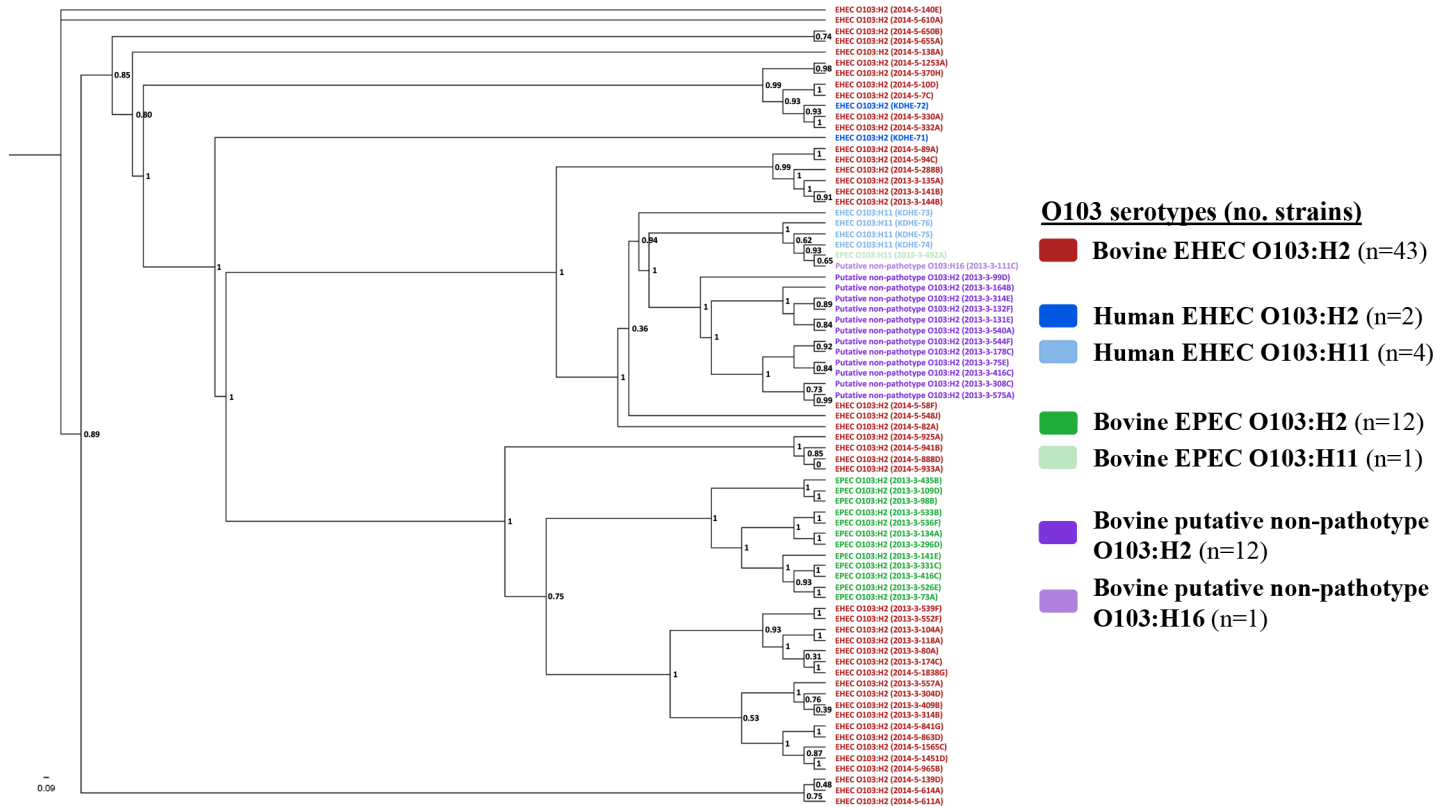
<sup>†</sup>Number of prophage sequences were determined from whole genome sequences of strains using Phage Search Tool Enhanced Release (PHASTER) [31, 32]. Only intact and questionable prophage counts based on PHASTER scores of >90 and 70–90, respectively, are shown.

<https://doi.org/10.1371/journal.pone.0191362.t005>

## Discussion

*Escherichia coli* O103 is the third most common STEC (next to O157 and O26) implicated in human STEC infections [1, 38]. Based on our studies, serogroup O103 is the second most prevalent STEC (next to O157) shed in cattle feces [12, 21]. Brooks *et al.* [38] have reported that 117 human clinical O103 isolates, submitted to CDC from 1983 to 2002, were positive for *stx1* and negative for *stx2*, and included only four flagellar types, H2, H11, H25 and non-motile. Similarly, all Shiga toxin-producing strains of cattle origin in this study (n = 43) were positive for *stx1* gene only, however, all possessed the H2 flagellar type. The predominance of the H2 flagellar type in bovine strains is in agreement with previous reports of O103 strains in cattle and sheep [20, 39–41]. The majority of EHEC strains (48/49; 98.0%) in our study had Shiga toxin 1a (*stx1a*) gene. Söderlund *et al.* [20] report Shiga toxin 1a (*stx1a*) subtype present in five EHEC O103:H2 isolated from Swedish cattle. Similar to findings from previous studies [20, 33], all EHEC/EPEC O103:H2 and O103:H11 strains carried epsilon and beta1 *ea*e subtypes, respectively. All EPEC strains included in this study were considered atypical, as indicated by the absence of the EAF plasmid, a finding also in agreement with previous studies [20, 42, 43].

All EHEC O103 strains in this study (43 bovine and 6 human strains) had a higher number of genes on mobile elements (prophages, transposable elements, and plasmids) compared to the bovine EPEC (except for one O103:H11 strain) and putative non-pathotype strains. Significant differences in the genome size observed among the O103 subgroups are reflective of the

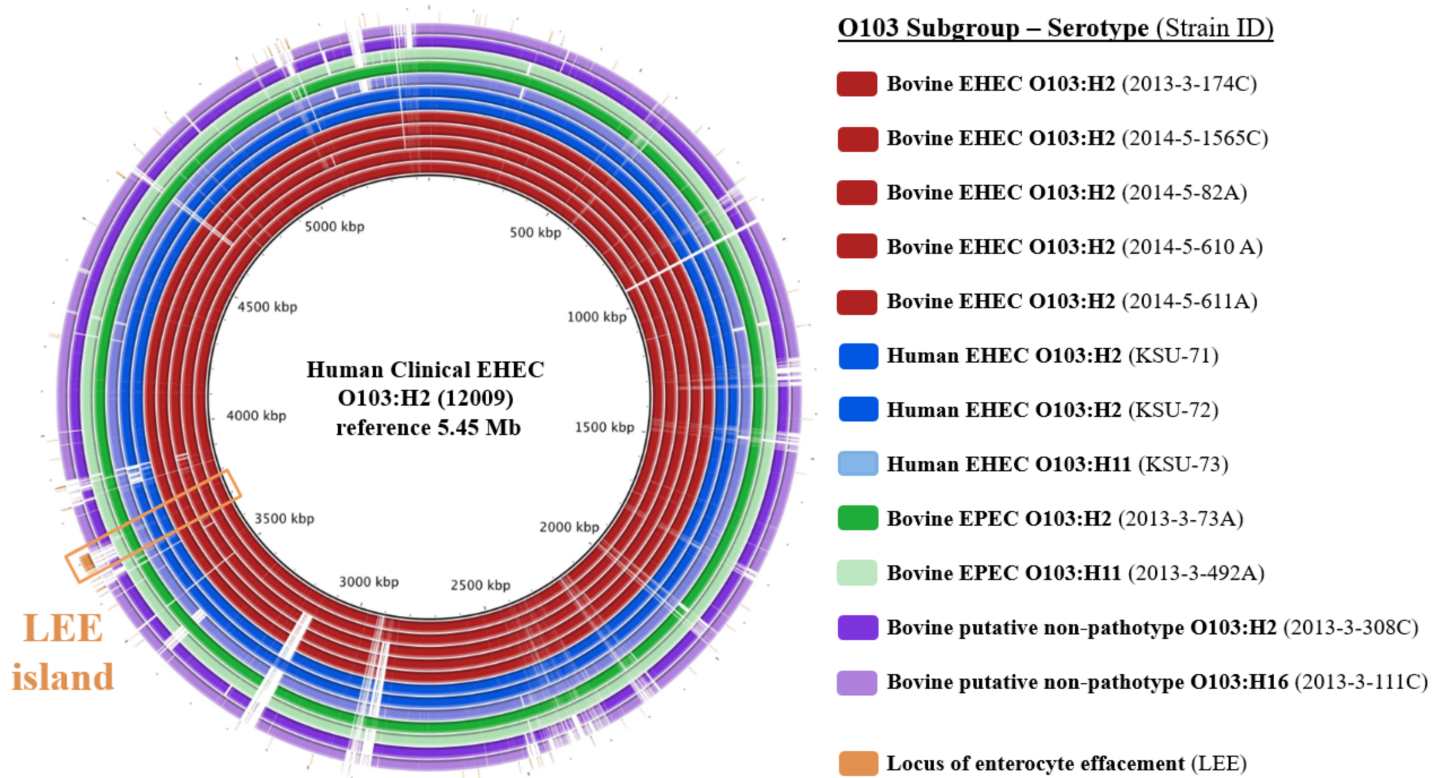


**Fig 2. Proportional branch transformed phylogenetic tree<sup>†</sup> of 75 strains of enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/ae* negative) *Escherichia coli* O103 of bovine and human origin using FigTree 1.4. <sup>†</sup>Numbers on the branches correspond to bootstrap values.**

<https://doi.org/10.1371/journal.pone.0191362.g002>

number of genes from mobile elements. However, one bovine EPEC O103:H11 strain (2013-3-492A) was an exception as its genome size and number of genes on mobile elements were more comparable to EHEC strains (Fig 1); furthermore, this strain was more closely related to a human EHEC O103:H11 strain (KSU-74) than to any of the EPEC strains (Fig 2). Also, the virulence gene profile of the EPEC O103:H11 strain 2013-3-492A more closely resembled the virulence gene profiles of the EHEC O103 subgroup than that of the bovine EPEC O103 subgroup. Furthermore, the strain is positive for *stx1* bacteriophage insertion site (*yehV*) and bacteriophage-*yehV* right and left junctions [44], suggesting that the EPEC O103:H11 strain may be capable of acquiring and/or had once acquired but lost *stx* gene(s). This suggests that much of the genetic diversity in *E. coli* O103 strains shed in cattle feces can be attributed to the loss or to acquisition of mobile genetic elements [45].

Similar to the phylogenetic clustering of bovine EHEC and EPEC O103:H2 strains reported in Söderlund *et al.* [20], strains in this study largely clustered by pathotype (Fig 2). A genome-wide visual comparison between representative strains from observed serotypes (O103:H2, O103:H11, O103:H16) within each O103 subgroup (bovine EHEC, human EHEC, bovine EPEC, and bovine putative non-pathotype) showed clear differences in the sequence identity between target strains (Fig 3). Interestingly, two of the bovine EHEC O103:H2 strains (2013-3-174C and 2014-5-1565C) shared more sequence identity with the clinical reference strain than did the human EHEC strains included in Fig 3, which may be an indication of the virulence potential of these strains. It is clear that the EHEC and EPEC strains have acquired more genetic elements during the course of their evolution in comparison to the putative non-



**Fig 3. Multiple genome comparison of representative strains of enterohemorrhagic (EHEC), enteropathogenic (EPEC) and putative non-pathotype (*stx/eae* negative) *Escherichia coli* O103 strains of bovine and human origin using BLAST Ring Image Generator (BRIG) v0.95.** †The nucleotide sequence (45,325 bp) of the locus of enterocyte effacement (LEE) pathogenicity island (GenBank accession no.: AF071034.1) was mapped for comparison of LEE between target strains.

<https://doi.org/10.1371/journal.pone.0191362.g003>

pathotype strains. Although overall number of genes implicated in virulence, disease and defense was comparable among all 69 bovine strains, a closer examination revealed key differences in virulence gene profiles of O103 subgroups and serotypes within subgroups.

### LEE effector genes

The chromosomal LEE pathogenicity island carries genes that encode for intimin (*eae*), translocated intimin receptor protein (*tir*), and type III secretion system effector proteins (*espA* and *espB*). Studies have shown that without any one of these genes (*eae*, *tir*, *espA*, *espB*), attaching and effacing (A/E) *E. coli* are unable to produce their characteristic A/E lesions [46–48]. The *espF* gene is also LEE encoded, but unlike the other LEE genes that were present in all EHEC and EPEC strains, a small number of bovine EPEC (3/13) and EHEC (4/43) strains were *espF*-negative. Although *espF* contributes to the disruption of intestinal barrier function during attachment, McNamara *et al.* [49] have shown that the gene is not required for A/E lesion formation. Other type III effector genes (*cif*, *espJ*), and *tccP*) were variably present in the EHEC and EPEC strains, possibly, because they are prophage-encoded genes. Although *cif* and *espJ* genes enhance attachment, *in vivo* and/or *in vitro* studies have shown that A/E lesions are not significantly inhibited in the absence of either gene [50, 51]. Garmendia *et al.* [52] have shown that *tir*-cytoskeleton coupling protein gene (*tccP*) assists in the translocation of the intimin receptor protein during bacterial attachment. In the same study, *tccP* mutants were unable to trigger A/E lesions on *in vitro*-inoculated HeLa epithelial cells. Considering its seemingly critical importance in type III secretory system-related disease outcomes, it is surprising that not

all human clinical EHEC were positive for the *tccP* gene. Garmendia *et al.* [52] reported that *tir* translocation was not affected in *tccP* mutants, therefore, it is possible that bacterial attachment and expression of other virulence factors in *tccP*-negative EHEC could contribute to A/E lesions.

### Non-LEE effector genes

Non-LEE effector (*nle*) genes, including *nleA*, *nleB* and *nleC*, have been associated with HUS-causing strains of EHEC [53] and were present in varying proportions within EHEC and EPEC O103 subgroups in this study. In two independent studies,  $\Delta nleA$  [54] and  $\Delta nleB$  mutant strains of *Citrobacter rodentium* [55] were unable to cause mortality in inoculated mice. Wickham *et al.* [55] also reported a three-log decrease ( $10^6$  vs.  $10^3$ ) in infectious dose for *nleB* wild-type- compared to  $\Delta nleB$ -mutant, which highlights the importance of *nleB* gene as it relates to the low infectious dose of EHEC strains. The *nleC* gene serves to down-regulate host NF- $\kappa$ B signaling pathway in efforts to disrupt immune clearance of invading bacteria [56]. Although *nleC* has also been significantly associated with HUS-causing strains [53], it was present only in 4 of 6 human clinical EHEC strains, but in 53.5% (23/43) of bovine EHEC strains.

### pO157 plasmid encoded virulence genes

The pO157 plasmid (~93 kb) carries a number of virulence genes implicated in EHEC virulence [57] and is present in nearly all clinical O157:H7 strains [58]. Major pO157 plasmid-encoded genes, *ehxA*, *espP*, *etpD*, *katP* and *toxB*, were present in many EHEC and EPEC O103 strains. The enterohemolysin gene (*ehxA*), present in all EHEC (49/49) and nearly all EPEC (12/13) strains in this study, encodes for a pore-forming toxin, which elicits *in vivo* production of IL-1 $\beta$  from human mononuclear cells, a commonly expressed cytokine during HUS infections [59]. The extracellular serine protease gene (*espP*) was found in almost all EHEC and EPEC strains and is considered to contribute to hemorrhagic colitis via the cleavage of pepsin A and human coagulation factor V [60].

The *etpD*, *katP* and *toxB* genes, located on the pO157 plasmid, were less frequently present in EHEC and EPEC strains, compared to *ehxA* and *espP* genes. Schmidt *et al.* [61] report that EHEC type II secretion pathway (*etp*) genes are not commonly detected (~10%) in bovine EHEC isolated from feces. In this study, *etpD* gene was present in 9 of 43 (20.9%) of bovine EHEC strains, but absent in the other subgroups. Brunder *et al.* [62] report a close association between the presence of *ehxA* and the catalase peroxidase gene (*katP*) in EHEC O157:H7 strains. We observed a similar trend for bovine and human EHEC; however, *ehxA* was present in a majority (11/12) of bovine EPEC O103:H2, whereas *katP* was absent in all of those strains. The *toxB* gene, identified by Tatsuno *et al.* [63], is a homolog of EHEC factor for adherence gene (*efa1*), carried on the pO157 plasmid and is commonly present in clinical EHEC O157:H7. In a study examining the prevalence of *toxB* in O157 and major non-O157 EHEC and EPEC of clinical and animal origin, Tozzoli *et al.* [64] report all O103 strains used in their study were negative for the gene. In the current study, 3 of 6 human EHEC strains were positive for *toxB*. Yet, the gene was present in only 1/43 bovine EHEC strains and in the single bovine EPEC O103:H11 strain. Although *toxB* is not required for formation of A/E lesions, Tatsuno *et al.* [63] showed that expression of *toxB* does lead to enhanced virulence by increasing expression of major LEE-encoded effector genes including *espA*, *espB* and *tir*.

### Other virulence genes

Interestingly, *lpfA* was the only adherence-based virulence gene present in the bovine putative non-pathotype O103:H2 strains (n = 12), yet the gene was absent in all EHEC (n = 43) and

EPEC O103:H2 (n = 12) strains, suggesting possible loss of *lpfA* gene by O103:H2 serotype at some point during the course of acquiring new genetic elements. The gene for increased serum survival (*iss*) was prevalent in all 75 strains. The *iss* gene is often associated with avian pathogenic *E. coli* (APEC) that cause colibacillosis in poultry, and serves as a genetic marker for APEC strains [4]. Among APEC, the *iss* gene is carried by a ColV plasmid [65] that in addition to conferring increased virulence and fitness traits, also encodes for multidrug resistance [66].

The *E. coli* secreted protease island encoded gene (*espI*) is considered part of the family of extracellular proteases known as SPATE, or serine protease autotransporters of *Enterobacteriaceae* [67]. The *espI* gene is harbored on the O91:H pathogenicity island and previously reported to occur exclusively in a LEE-negative subgroup of STEC that carry a *stx2d* gene variant [68]. Krüger *et al.* [69] also report detection of *espI* gene exclusively in *stx2*- (but not *stx1*) positive *E. coli* O26:H11 strains of clinical, bovine and food origin. In our study, *espI* gene was present in more than half (23/43; 53.5%) of all bovine EHEC O103:H2 that were *stx1a* positive; *espI* gene was also present in three of 12 bovine EPEC O103:H2 strains. These results are in contrast with previous studies linking the *espI* gene to *stx2*-carrying EHEC only [68, 69] and may be the first time *espI* gene has been reported in bovine EHEC and EPEC O103 strains.

### Plasmid and prophage sequences

Some of these plasmid sequences are originally associated with non-*E. coli* bacteria, including *Klebsiella pneumoniae* (ColRNAI and IncA/C2), *Salmonella typhi* (IncFIA(HI1)), *Salmonella typhimurium* (IncN) and *Pseudomonas aeruginosa* (IncP), which further highlights the mobility of these genetic elements. Many of the plasmids, including IncA/C2, IncFII, IncFII (pHN7A8), IncFII(pRSB107), IncN and IncX1 have also been associated with antimicrobial resistance determinants and/or other putative virulence-associated functions, that in some cases have been the causative genetic element behind human outbreaks [70]. The IncF incompatibility family represents the majority of virulence-associated plasmids carried by *E. coli* [71], therefore it may not be surprising that IncF plasmids represented nearly half (96/218; 44.0%) of all total plasmids identified in the strains used in this study.

Similarly, non-*E. coli* prophage sequences, including *Aeromonas* phage phiO18P, *Burkholderia* phage phiE255, *Salmonella* phage SEN34 and *Shigella* phage SflI, were found in many of the strains, which further demonstrates the mobility of these genetic elements. The most and least prophage diversity, defined by total number of different prophages carried by an O103 subgroup, was found in bovine EHEC and bovine putative non-pathotype strains, respectively, which also highlights the differences in mobile content found between these subgroups.

### Conclusion

The virulence gene profiles of the bovine and human EHEC, bovine (atypical) EPEC and putative non-pathotype strains of *E. coli* O103 were quite diverse. The difference in the number of strains tested within each subgroup and lack of publicly available O103 genome sequences may have limited the strength of comparison. Although the *in silico* analysis performed here does not provide phenotypic evidence of virulence contributions, a number of major and putative virulence genes were comparable among bovine and human EHEC O103 strains, which may indicate the potential for bovine EHEC O103 to cause human infection. The bovine EPEC O103:H11 strain also shared similar virulence gene and plasmid profiles with human EHEC O103:H11 strains, raising the possibility that the EPEC may have lost its *stx* prophage. Regardless, the *in silico* data highlight the numerous virulence genes carried on mobile genetic elements (prophages, transposable elements, and plasmids) that contribute to the plasticity of



bovine EHEC or EPEC. Genome size and number of genes from mobile elements were strongly correlated among the O103 subgroups. The putative non-pathotype strains had the smallest genome size and carried the fewest overall number of mobile genes and perhaps related to this, lacked any specific major or putative mobile virulence genes. The EPEC strains in this study had larger genomes and were positive for a higher number of specific virulence genes compared to putative non-pathotype strains. Excluding the outlying EPEC O103:H11 strain, the EHEC overshadowed EPEC, and putative non-pathotype subgroups in both these categories, which raises the question whether progenitor EHEC bacteria are more genetically predisposed toward acquiring certain mobile elements that could confer virulence. Conversely, putative virulence genes that allow for increased EHEC survival within the environment or within a host may afford EHEC with increased opportunity to acquire mobile genetic elements. We believe that the diversity of pathotypes of *E. coli* O103 harbored and shed in the feces of cattle is reflective of the loss or acquisition of genes carried on mobile genetic elements. The environmental and biological mechanisms that allow for loss or acquisition of virulence genes by EHEC and EPEC and putative non-pathotype strains remain an exciting frontier for the whole-genome sequence-based analysis of *E. coli* pathotypes.

## Supporting information

**S1 Table. Virulence gene profiles<sup>†</sup> of enterohemorrhagic *Escherichia coli* (EHEC) O103:H2 strains isolated from cattle feces collected from nine feedlots in the Midwest.** <sup>†</sup>Virulence genes were determined using Virulence Finder 1.4 [29].  
(DOCX)

**S2 Table. Virulence gene profiles<sup>†</sup> of enteropathogenic *Escherichia coli* (EPEC) O103 and *E. coli* O103 strains negative for Shiga toxin and intimin genes (O-group) isolated from cattle feces collected from a Midwest feedlot.** <sup>†</sup>Virulence genes were determined using Virulence Finder 1.4 [29].  
(DOCX)

**S3 Table. Virulence gene profiles<sup>†</sup> of clinical human enterohemorrhagic *Escherichia coli* (EHEC) O103 strains.** <sup>†</sup>Virulence genes were determined using Virulence Finder 1.4 [29].  
<sup>‡</sup>Control strains were included for comparison and result from the testing of genomic and plasmid (O103:H2 12009, NC\_013354.1; Sakai, NC\_002128.1 and NC\_002127.1; EDL933, AF074613.1) DNA sequences available at GenBank.  
(DOCX)

**S4 Table. Plasmid profiles<sup>†</sup> of enterohemorrhagic *Escherichia coli* (EHEC) O103:H2 strains isolated from cattle feces collected from nine feedlots in the Midwest.** <sup>†</sup>Plasmids were determined from whole genome sequences of strains using Plasmid Finder 1.3 [30].  
(DOCX)

**S5 Table. Plasmid profiles<sup>†</sup> of enteropathogenic *Escherichia coli* (EPEC) O103 and *E. coli* O103 strains negative for Shiga toxin and intimin genes (O-group) isolated from cattle feces collected from a Midwest feedlot.** <sup>†</sup>Plasmids were determined from whole genome sequences of strains using Plasmid Finder 1.3 [30].  
(DOCX)

**S6 Table. Plasmid profiles<sup>†</sup> of clinical human enterohemorrhagic *Escherichia coli* (EHEC) O103 strains.** <sup>†</sup>Plasmids were determined from whole genome sequences of strains using Plasmid Finder 1.3 [30].

\*Control strains were included for comparison and result from the testing of genomic and

plasmid (O103:H2 12009, NC\_013354.1; Sakai, NC\_002128.1 and NC\_002127.1; EDL933, AF074613.1) DNA sequences available at GenBank.  
(DOCX)

**S7 Table. Prophage profiles<sup>†</sup> of enterohemorrhagic *Escherichia coli* (EHEC) O103:H2 strains isolated from cattle feces collected from nine feedlots in the Midwest.** <sup>†</sup>Prophage sequences were determined from whole genome sequences of strains using Phage Search Tool Enhanced Release (PHASTER) [31, 32]. Only intact and questionable prophage counts based on PHASTER scores of >90 and 70–90, respectively, are shown.  
(DOCX)

**S8 Table. Prophage profiles<sup>†</sup> of enteropathogenic *Escherichia coli* (EPEC) O103 and *E. coli* O103 strains negative for Shiga toxin and intimin genes (O-group) isolated from cattle feces collected from a Midwest feedlot.** <sup>†</sup>Prophage sequences were determined from whole genome sequences of strains using Phage Search Tool Enhanced Release (PHASTER) [31, 32]. Only intact and questionable prophage counts based on PHASTER scores of >90 and 70–90, respectively, are shown.  
(DOCX)

**S9 Table. Prophage profiles<sup>†</sup> of clinical human enterohemorrhagic *Escherichia coli* (EHEC) O103 strains.** <sup>†</sup>Prophage sequences were determined from whole genome sequences of strains using Phage Search Tool Enhanced Release (PHASTER) [31, 32]. Only intact and questionable prophage counts based on PHASTER scores of >90 and 70–90, respectively, are shown.  
\*Control strains were included for comparison and result from the testing of genomic and plasmid (O103:H2 12009, NC\_013354.1; Sakai, NC\_002128.1 and NC\_002127.1; EDL933, AF074613.1) DNA sequences available at GenBank.  
(DOCX)

## Acknowledgments

Contribution no.17-244-J from the Kansas Agricultural Experiment Station. The authors wish to thank Neil Wallace for his assistance in this project.

## Author Contributions

**Conceptualization:** T. G. Nagaraja.

**Data curation:** Jay N. Worley, Xun Yang, Doina Caragea.

**Formal analysis:** Jay N. Worley, Xun Yang, Justin B. Ludwig, Doina Caragea.

**Funding acquisition:** Jianfa Bai, T. G. Nagaraja.

**Investigation:** Lance W. Noll, Jay N. Worley, Xun Yang, Pragathi B. Shridhar, Justin B. Ludwig, Xiaorong Shi, Jianfa Bai, T. G. Nagaraja.

**Methodology:** Pragathi B. Shridhar, Xiaorong Shi, Jianfa Bai, T. G. Nagaraja.

**Project administration:** Jianfa Bai, Jianghong Meng, T. G. Nagaraja.

**Resources:** T. G. Nagaraja.

**Software:** T. G. Nagaraja.

**Supervision:** T. G. Nagaraja.

**Writing – original draft:** Lance W. Noll, T. G. Nagaraja.

Writing – review & editing: Jianfa Bai.

## References

1. Centers for Disease Control and Prevention (CDC). Shiga toxin-producing *Escherichia coli* (STEC) surveillance annual summary, 2012. Atlanta, Georgia: US Department of Health and Human Services, CDC, 2014.
2. Abdullah UY, Al-Sultan II, Jassim HM, Ali YA, Khorsheed RM, Baig AA. Hemolytic uremic syndrome caused by Shiga toxin-producing *Escherichia coli* infections: an overview. *Cloning & Transgenesis*. 2014; 3: 1–9.
3. Karmali MA. Infection by verocytotoxin-producing *Escherichia coli*. *Clinical Microbiol Rev*. 1989; 2: 15–38.
4. Ewers C, Janßen T, Wieler L. Avian pathogenic *Escherichia coli* (APEC). *Berliner und Munchener tierärztliche Wochenschrift*. 2002; 116: 381–395.
5. Jordan DM, Cornick N, Torres AG, Dean-Nystrom EA, Kaper JB, Moon HW. Long polar fimbriae contribute to colonization by *Escherichia coli* O157: H7 in vivo. *Infect Immun*. 2004; 72: 6168–6171. <https://doi.org/10.1128/IAI.72.10.6168-6171.2004> PMID: 15385526
6. Stevens MP, Roe AJ, Vlisidou I, Van Diemen PM, La Ragione RM, Best A, et al. Mutation of toxB and a truncated version of the efa-1 gene in *Escherichia coli* O157: H7 influences the expression and secretion of locus of enterocyte effacement-encoded proteins but not intestinal colonization in calves or sheep. *Infect Immun*. 2004; 72: 5402–5411. <https://doi.org/10.1128/IAI.72.9.5402-5411.2004> PMID: 15322038
7. Trabulsi LR, Keller R, Gomes TAT. Typical and atypical enteropathogenic *Escherichia coli*—Synopsis. *Emerg Infect Dis*. 2002; 8: 508–514. <https://doi.org/10.3201/eid0805.010385> PMID: 11996687
8. Donnenberg MS, Finlay BB. Combating enteropathogenic *Escherichia coli* (EPEC) infections: the way forward. *Trends Microbiol*. 2013; 21: 317–319. <https://doi.org/10.1016/j.tim.2013.05.003> PMID: 23815982
9. Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11:142–201. PMID: 9457432
10. Mellmann A, Lu S, Karch H, Xu J-g, Harmsen D, Schmidt MA, et al. Recycling of Shiga toxin 2 genes in sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157: NM. *Appl Environ Microb*. 2008; 74: 67–72.
11. Bielaszewska M, Middendorf B, Friedrich AW, Fruth A, Karch H, Schmidt MA, et al. Shiga toxin-negative attaching and effacing *Escherichia coli*: distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. *Clin Infect Dis*. 2008; 47: 208–217. <https://doi.org/10.1086/589245> PMID: 18564929
12. Noll LW, Shridhar PB, Dewsbury DM, Shi X, Cernicchiaro N, Renter DG, et al. A comparison of culture- and PCR-based methods to detect six major non-O157 serogroups of Shiga toxin-producing *Escherichia coli* in cattle feces. *PloS One*. 2015 Aug 13. e0135446. <https://doi.org/10.1371/journal.pone.0135446>
13. Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, et al. Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. *Nature*. 2001; 409: 529–533. <https://doi.org/10.1038/35054089> PMID: 11206551
14. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *PNAS*. 2009; 106: 17939–17944. <https://doi.org/10.1073/pnas.0903585106> PMID: 19815525
15. Ison SA, Delannoy S, Bugarel M, Nightingale KK, Webb HE, Renter DG, et al. Genetic diversity and pathogenic potential of attaching and effacing *Escherichia coli* O26: H11 strains recovered from bovine feces in the United States. *Appl Environ Microbiol*. 2015; 81: 3671–3678. <https://doi.org/10.1128/AEM.00397-15> PMID: 25795673
16. Norman KN, Clawson ML, Strockbine NA, Mandrell RE, Johnson R, Ziebell K, et al. Comparison of whole genome sequences from human and non-human *Escherichia coli* O26 strains. *Front Cell Infect Microbiol*. 2015; 5: 1–10. <https://doi.org/10.3389/fcimb.2015.00001>
17. Gonzalez-Escalona N, Toro M, Rump LV, Cao G, Nagaraja TG, Meng J. Virulence gene profiles and clonal relationships of *Escherichia coli* O26: H11 isolates from feedlot cattle by whole genome sequencing. *Appl Environ Microbiol*. 2016; 82: 3900–3912. <https://doi.org/10.1128/AEM.00498-16> PMID: 27107118
18. Carter MQ, Quinones B, He X, Zhong W, Louie JW, Lee BG, et al. Clonal population of environmental Shiga toxin-producing *Escherichia coli* O145 exhibits large phenotypic variation including virulence

- traits. *Appl Environ Microbiol.* 2015; 82: 1090–1101. <https://doi.org/10.1128/AEM.03172-15> PMID: 26637597
19. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157: H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 2001; 8: 11–22. PMID: 11258796
  20. Söderlund R, Hurel J, Jinnerot T, Sekse C, Aspán A, Eriksson E, et al. Genomic comparison of *Escherichia coli* serotype O103: H2 isolates with and without verotoxin genes: implications for risk assessment of strains commonly found in ruminant reservoirs. *Infect Ecol Epidemiol.* 2016; 6: 1–6.
  21. Dewsbury DM, Renter DG, Shridhar PB, Noll LW, Shi X, Nagaraja TG, et al. Summer and winter prevalence of Shiga toxin-producing *Escherichia coli* (STEC) O26, O45, O103, O111, O121, O145, and O157 in feces of feedlot cattle. *Foodborne Pathog Dis.* 2015; 12: 726–732. <https://doi.org/10.1089/fpd.2015.1987> PMID: 26075548
  22. Cull CA, Renter DG, Dewsbury DM, Noll LW, Shridhar PB, Ives SE, et al. Feedlot- and pen-level prevalence of enterohemorrhagic *Escherichia coli* in feces of commercial feedlot cattle in two major U.S. cattle feeding areas. *Foodborne Pathog Dis.* 2017; 14: 309–317. <https://doi.org/10.1089/fpd.2016.2227> PMID: 28281781
  23. Bai J, Paddock ZD, Shi X, Li S, An B, Nagaraja TG. Applicability of a multiplex PCR to detect the seven major Shiga toxin-producing *Escherichia coli* based on genes that code for serogroup-specific O-antigens and major virulence factors in cattle feces. *Foodborne Pathog Dis.* 2012; 9: 541–548. <https://doi.org/10.1089/fpd.2011.1082> PMID: 22568751
  24. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012; 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021> PMID: 22506599
  25. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008; 9: 1–15. <https://doi.org/10.1186/1471-2164-9-1>
  26. Kwon T, Kim J-B, Bak Y-S, Yu Y-B, Kwon KS, Kim W, et al. Draft genome sequence of non-Shiga toxin-producing *Escherichia coli* O157 NCCP15738. *Gut Pathog.* 2016; 8: 1. <https://doi.org/10.1186/s13099-015-0083-z>
  27. Ferdous M, Zhou K, Mellmann A, Morabito S, Croughs PD, de Boer RF, et al. Is Shiga toxin-negative *Escherichia coli* O157: H7 enteropathogenic or enterohemorrhagic *Escherichia coli*? Comprehensive molecular analysis using whole-genome sequencing. *J Clin Microbiol.* 2015; 53: 3530–3538. <https://doi.org/10.1128/JCM.01899-15> PMID: 26311863
  28. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J Clin Microbiol.* 2015; 53: 2410–2426. <https://doi.org/10.1128/JCM.00008-15> PMID: 25972421
  29. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol.* 2014; 52: 1501–1510. <https://doi.org/10.1128/JCM.03617-13> PMID: 24574290
  30. Carattoli A, Zankari E, Garcia-Fernandez A, Larsen MV, Lund O, Villa L, et al. PlasmidFinder and pMLST: in silico detection and typing of plasmids. *Antimicrob Agents Ch.* 2014; ACC: 02412–02414.
  31. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res.* 2011; 39: W347–W352. <https://doi.org/10.1093/nar/gkr485> PMID: 21672955
  32. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 2016; 44: W16–W21. <https://doi.org/10.1093/nar/gkw387> PMID: 27141966
  33. Iguchi A, Iyoda S, Ohnishi M. Molecular characterization reveals three distinct clonal groups among clinical Shiga toxin-producing *Escherichia coli* strains of serogroup O103. *J Clin Microbiol.* 2012; 50: 2894–2900. <https://doi.org/10.1128/JCM.00789-12> PMID: 22718945
  34. Nadya S, Delaquis P, Chen J, Allen K, Johnson RP, Ziebell K, et al. Phenotypic and genotypic characteristics of Shiga toxin-producing *Escherichia coli* isolated from surface waters and sediments in a Canadian urban-agricultural landscape. *Front Cell Infect Microbiol.* 2016; 6: 1–13. <https://doi.org/10.3389/fcimb.2016.00001>
  35. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 2014; 15: 1–15.
  36. Rambaut A. FigTree v. 1.4. 2. 2014.
  37. Alikhan N-F, Petty NK, Zakour NLB, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* 2011; 12: 1–10.

38. Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, et al. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. *J Infectious Dis.* 2005; 192: 1422–1429.
39. Blanco M, Blanco J, Mora A, Dahbi G, Alonso M, González E, et al. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from cattle in Spain and identification of a new intimin variant gene (*eae-ξ*). *J Clin Microbiol.* 2004; 42: 645–651. <https://doi.org/10.1128/JCM.42.2.645-651.2004> PMID: 14766831
40. Padola NL, Sanz ME, Blanco JE, Blanco M, Blanco J, Etcheverria AI, et al. Serotypes and virulence genes of bovine Shigatoxigenic *Escherichia coli* (STEC) isolated from a feedlot in Argentina. *Vet Microbiol.* 2004; 100: 3–9. [https://doi.org/10.1016/S0378-1135\(03\)00127-5](https://doi.org/10.1016/S0378-1135(03)00127-5) PMID: 15135507
41. Sekse C, Sunde M, Hopp P, Bruheim T, Cudjoe KS, Kvittle B, et al. Occurrence of potentially human-pathogenic *Escherichia coli* O103 in Norwegian sheep. *Appl Environ Microbiol.* 2013; 79: 7502–7509. <https://doi.org/10.1128/AEM.01825-13> PMID: 24077709
42. Sandhu K, Clarke R, Gyles C. Virulence markers in Shiga toxin-producing *Escherichia coli* isolated from cattle. *Can J Vet Res.* 1999; 63: 177–184. PMID: 10480459
43. Paddock ZD, Renter DG, Cull CA, Shi X, Bai J, Nagaraja TG. *Escherichia coli* O26 in feedlot cattle: fecal prevalence, isolation, characterization, and effects of an *E. coli* O157 vaccine and a direct-fed microbial. *Foodborne Pathog Dis.* 2014; 11: 186–193. <https://doi.org/10.1089/fpd.2013.1659> PMID: 24286301
44. Shaikh N, Tarr PI. *Escherichia coli* O157: H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J Bacteriol.* 2003; 185: 3596–3605. <https://doi.org/10.1128/JB.185.12.3596-3605.2003> PMID: 12775697
45. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature.* 2000; 405: 299–304. <https://doi.org/10.1038/35012500> PMID: 10830951
46. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *PNAS.* 1995; 92: 1664–1668. PMID: 7878036
47. McDaniel TK, Kaper JB. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Molec Microbiol.* 1997; 23: 399–407.
48. Abe A, Heczko U, Hegele RG, Finlay BB. Two enteropathogenic *Escherichia coli* type III secreted proteins, EspA and EspB, are virulence factors. *J Exp Med.* 1998; 188: 1907–1916. PMID: 9815268
49. McNamara BP, Koutsouris A, O’Connell CB, Nougayrède J-P, Donnenberg MS, Hecht G. Translocated EspF protein from enteropathogenic *Escherichia coli* disrupts host intestinal barrier function. *J Clin Invest.* 2001; 107: 621–629. <https://doi.org/10.1172/JCI11138> PMID: 11238563
50. Dahan S, Wiles S, La Ragione RM, Best A, Woodward MJ, Stevens MP, et al. EspJ is a prophage-carried type III effector protein of attaching and effacing pathogens that modulates infection dynamics. *Infect Immun.* 2005; 73: 679–686. <https://doi.org/10.1128/IAI.73.2.679-686.2005> PMID: 15664905
51. Marchès O, Ledger TN, Boury M, Ohara M, Tu X, Goffaux F, et al. Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G2/M transition. *Molec Microbiol.* 2003; 50: 1553–1567.
52. Garmendia J, Phillips AD, Carlier MF, Chong Y, Schüller S, Marches O, et al. TccP is an enterohaemorrhagic *Escherichia coli* O157: H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol.* 2004; 6: 1167–1183. <https://doi.org/10.1111/j.1462-5822.2004.00459.x> PMID: 15527496
53. Bugarel M, Martin A, Fach P, Beutin L. Virulence gene profiling of enterohaemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis for molecular risk assessment of typical and atypical EPEC strains. *BMC Microbiol.* 2011; 11: 1–10. <https://doi.org/10.1186/1471-2180-11-1>
54. Gruenheid S, Sekirov I, Thomas NA, Deng W, O’Donnell P, Goode D, et al. Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic *Escherichia coli* O157: H7. *Molec Microbiol.* 2004; 51: 1233–1249.
55. Wickham ME, Lupp C, Mascarenhas M, Vázquez A, Coombes BK, Brown NF, et al. Bacterial genetic determinants of non-O157 STEC outbreaks and hemolytic-uremic syndrome after infection. *J Infect Dis.* 2006; 194: 819–827. <https://doi.org/10.1086/506620> PMID: 16941350
56. Yen H, Ooka T, Iguchi A, Hayashi T, Sugimoto N, Tobe T. NleC, a type III secretion protease, compromises NF-κB activation by targeting p65/RelA. *PLoS Pathog.* 2010 Dec 16. e1001231. <https://doi.org/10.1371/journal.ppat.1001231> PMID: 21187904
57. Makino K, Ishii K, Yasunaga T, Hattori M, Yokoyama K, Yutsudo CH, et al. Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohaemorrhagic *Escherichia coli* O157: H7 derived from Sakai outbreak. *DNA Res.* 1998; 5: 1–9. PMID: 9628576
58. Schmidt H, Kernbach C, Karch H. Analysis of the EHEC hly operon and its location in the physical map of the large plasmid of enterohaemorrhagic *Escherichia coli* O157: H7. *Microbiol.* 1996; 142: 907–914.



59. Taneike I, Zhang H-M, Wakisaka-Saito N, Yamamoto T. Enterohemolysin operon of Shiga toxin-producing *Escherichia coli*: a virulence function of inflammatory cytokine production from human monocytes. *FEBS Lett.* 2002; 524: 219–224. PMID: [12135770](#)
60. Brunder W, Schmidt H, Karch H. EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157: H7 cleaves human coagulation factor V. *Mol Microbiol.* 1997; 24: 767–778. PMID: [9194704](#)
61. Schmidt H, Henkel B, Karch H. A gene cluster closely related to type II secretion pathway operons of gram-negative bacteria is located on the large plasmid of enterohaemorrhagic *Escherichia coli* O157 strains. *FEMS Microbiol Lett.* 1997; 148: 265–272. PMID: [9084155](#)
62. Brunder W, Schmidt H, Karch H. KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157: H7. *Microbiol.* 1996; 142: 3305–3315
63. Tatsuno I, Horie M, Abe H, Miki T, Makino K, Shinagawa H, et al. *tox*B gene on pO157 of enterohaemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. *Infect Immun.* 2001; 69: 6660–6669. <https://doi.org/10.1128/IAI.69.11.6660-6669.2001> PMID: [11598035](#)
64. Tozzoli R, Caprioli A, Morabito S. Detection of *tox*B, a plasmid virulence gene of *Escherichia coli* O157, in enterohaemorrhagic and enteropathogenic *E. coli*. *J Clin Microbiol.* 2005; 43: 4052–4056. <https://doi.org/10.1128/JCM.43.8.4052-4056.2005> PMID: [16081950](#)
65. Johnson TJ, Siek KE, Johnson SJ, Nolan LK. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J Bacteriol.* 2006; 188: 745–758. <https://doi.org/10.1128/JB.188.2.745-758.2006> PMID: [16385064](#)
66. Johnson TJ, Wannemuehler YM, Nolan LK. Evolution of the *iss* gene in *Escherichia coli*. *Appl Environ Microbiol.* 2008; 74: 2360–2369. <https://doi.org/10.1128/AEM.02634-07> PMID: [18281426](#)
67. Dautin N. Serine protease autotransporters of enterobacteriaceae (SPATEs): biogenesis and function. *Toxins.* 2010; 2: 1179–1206. <https://doi.org/10.3390/toxins2061179> PMID: [22069633](#)
68. Schmidt H, Zhang W-L, Hemmrich U, Jelacic S, Brunder W, Tarr P, et al. Identification and characterization of a novel genomic island integrated at *selC* in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. *Infect Immun.* 2001; 69: 6863–6873. <https://doi.org/10.1128/IAI.69.11.6863-6873.2001> PMID: [11598060](#)
69. Krüger A, Lucchesi PM, Sanso AM, Etcheverría AI, Bustamante AV, Burgán J, et al. Genetic characterization of Shiga toxin-producing *Escherichia coli* O26: H11 strains isolated from animal, food, and clinical samples. *Front Cell Infect Microbiol.* 2015; 5: 1–8. <https://doi.org/10.3389/fcimb.2015.00001>
70. Boyd DA, Tyler S, Christianson S, McGeer A, Muller MP, Willey BM, et al. Complete nucleotide sequence of a 92-kilobase plasmid harboring the CTX-M-15 extended-spectrum beta-lactamase involved in an outbreak in long-term-care facilities in Toronto, Canada. *Antimicrob Agents Ch.* 2004; 48: 3758–3764.
71. Johnson TJ, Nolan LK. Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol Mol Biol Rev.* 2009; 73: 750–774. <https://doi.org/10.1128/MMBR.00015-09> PMID: [19946140](#)