Genome wide characterization of enterotoxigenic *Escherichia coli* serogroup O6 isolates from multiple outbreaks and sporadic infections from 1975-2016

Vaishnavi Pattabiraman[ORCID][1]*, **Lee S. Katz**[1,3], **Jessica C. Chen**[1], **Andre E. McCullough**[2], **Eija Trees**[1]

1 Centers for Disease Control and Prevention, Atlanta, GA, United States of America, 2 IHRC Inc, Atlanta, GA, United States of America, 3 Center for Food Safety, College of Agricultural and Environmental Sciences, University of Georgia, Griffin, GA, United States of America

* Vxe9@cdc.gov

Abstract

Enterotoxigenic *Escherichia coli* (ETEC) are an important cause of diarrhea globally, particularly among children under the age of five in developing countries. ETEC O6 is the most common ETEC serogroup, yet the genome wide population structure of isolates of this serogroup is yet to be determined. In this study, we have characterized 40 ETEC O6 isolates collected between 1975–2016 by whole genome sequencing (WGS) and by phenotypic antimicrobial susceptibility testing. To determine the relatedness of isolates, we evaluated two methods—whole genome high-quality single nucleotide polymorphism (whole genome-hqSNP) and core genome SNP analyses using Lyve-SET and Parsnp respectively. All isolates were tested for antimicrobial susceptibility using a panel of 14 antibiotics. ResFinder 2.1 and a custom quinolone resistance determinants workflow were used for resistance determinant detection. VirulenceFinder 1.5 was used for prediction of the virulence genes. Thirty-seven isolates clustered into three major clades (I, II, III) by whole genome-hqSNP and core genome SNP analyses, while three isolates included in the whole genome-hqSNP analysis only did not cluster with clades I-III by both analyses and formed a distantly related outgroup, designated clade IV. Median number of pairwise whole genome-hqSNPs in clonal ETEC O6 outbreaks ranged from 0 to 5. Of the 40 isolates tested for antimicrobial susceptibility, 18 isolates were pansusceptible. Twenty-two isolates were resistant to at least one antibiotic, nine of which were multidrug resistant. Phenotypic antimicrobial resistance (AR) correlated with AR determinants in 22 isolates. Thirty-two isolates harbored both enterotoxin virulence genes while the remaining 8 isolates had only one of the two virulence genes. In summary, whole genome-hqSNP and core genome SNP analyses from this study revealed similar evolutionary relationships and an overall diversity of ETEC O6 isolates independent of time of isolation. Less than 5 pairwise hqSNPs between ETEC O6 isolates is circumstantially indicative of an outbreak cluster. Findings from this study will be a basis for quicker outbreak detection and control by efficient subtyping by WGS.

Introduction

Enterotoxigenic *Escherichia coli* (ETEC) is the leading bacterial cause of foodborne illnesses worldwide [1] and is the most common cause of bacterial diarrhea in children under the age of five in low and middle-income countries and in travelers to endemic areas [2, 3]. ETEC strains secrete heat-labile (LT) and/or heat-stable (ST) enterotoxins and produce colonization factors, which are fimbrial, afimbrial or fibrillar surface structures that mediate adherence to the human intestinal mucosa. For disease presentation, the enterotoxins must be successfully delivered to cognate receptors on epithelial cells of the small intestine, where ensuing loss of water and electrolytes results in diarrhea [4]. ETEC strains are genetically diverse and over a 100 different O antigens have been associated with clinical ETEC isolates [5–7] with O6 being the most common serogroup, both in the frequency of isolates and the number of geographic locations from where it was recovered [6, 8]. Both LT and ST have been reported in O6 strains [9, 10].

In upper-middle income countries such as Peru, fluoroquinolones (ciprofloxacin or norfloxacin) and azithromycin remain the drugs of choice for ETEC infection [11] whereas in lower-income countries such as Bangladesh high levels of antimicrobial resistance (>40%) to multiple individual antibiotics have been observed in addition to resistance to treatment options such as fluoroquinolones [12]. Centers for Disease Control and Prevention (CDC) indicate that fluoroquinolones are shown to be an effective therapy in ETEC mediated traveler's diarrhea [13]. Collectively, these studies connote that while fluoroquinolones remain the most common drug of choice, effective antibiotic treatment for ETEC mediated diarrhea may vary slightly depending on the geographical location.

A few comparative genomics studies of ETEC have been conducted that have shown substantial genetic diversity in this pathotype [10, 14–16]. For instance, von Mentzer *et al* [14] used next-generation sequencing (NGS) for a complete understanding of ETEC phylogeny and evolution using a global collection of ETEC isolates collected between 1980 and 2011 while Sahl *et al* [15] characterized phylogenomic diversity through the identification of genetically distinct pathogenic isolates within an ETEC infected individual by genome sequencing and comparative analysis.

However, no genome wide population studies were conducted on ETEC O6 isolates which is the most common ETEC serogoup. Here, we use NGS and antimicrobial susceptibility testing to characterize ETEC O6 isolates from multiple outbreaks and sporadic infections from 1975–2016. We describe the phylogeny and evolution of O6 isolates by whole genome high quality single nucleotide polymorphism (WG-hqSNP) and core genome SNP (CG SNP) analyses. We document the nature of antimicrobial resistance by phenotypic and genotypic antimicrobial resistance determination in these isolates. Understanding the genetic structure of the population as well as variation among strains that are likely to be epidemiologically related or unrelated will be helpful as many surveillance networks, including PulseNet USA, the molecular network for foodborne disease surveillance, moves forward with implementation of whole genome sequencing (WGS) and establishment of interpretive guidance.

Materials and methods

Bacterial strains

To characterize historic and recent ETEC serogroup O6 strains by WGS we selected 40 ETEC O6 isolates [38 O6:H16; 2 O6: non-motile (NM)] from 1975–2016 for inclusion in our study. At least one isolate was selected from each of the 18 outbreaks and 13 sporadic infections for further characterization. The isolates belonged to the Centers for Disease Control and

Prevention (Atlanta, GA) isolate collection. ETEC O6:H16 serotypes were predicted by SerotypeFinder-1.1 in the Center for Genomic Epidemiology (CGE) [17] website while O6:NM serotypes were confirmed by standard serotyping laboratory procedure [18]. These isolates are listed in S1 Table.

Total DNA extraction and quantification

Total DNA were extracted from the ETEC O6 isolates (S1 Table) using the DNeasy Blood and Tissue Kit (Qiagen Inc) as described in the Laboratory Standard Operating Procedure (SOP) for PulseNet Nextera XT library preparation and run set up for Illumina MiSeq [19]. Purity of DNA at 260/280 nm was confirmed by Nanodrop 2000 spectrophotometer (Thermo Scientific) and concentration of extracted DNA was determined by a Qubit 2.0 fluorometer (Life Technologies) using dsDNA BR assay kit. DNA that passed the quality control with a purity of >/ = 1.75 at 260/280 nm and with a minimum concentration of 10 ng/μl were subjected to NGS by MiSeq (Illumina Inc, San Diego, CA).

Whole genome sequencing

Sequencing libraries for 40 ETEC O6 isolates (S1 Table) were prepared by Nextera XT DNA library preparation kit as described [19]. Denatured pooled libraries were loaded at a final concentration of 10 pM on the flow cell. 500-cycle chemistry (2 x 250 bp reads) was used for sequencing the genomes of all isolates in this study. The PulseNet-recommended minimum quality thresholds of $\geq$ 40x coverage and average Q30 score $\geq$ 30 were followed for all sequenced genomes in this study [20]. The genome of the reference isolate 2011EL1370-2 (S1 Table) was sequenced using PacBio (Pacific Biosciences, Menlo Park, CA) as previously described [21] and the sequence reads were filtered and assembled *de novo* using the PacBio Hierarchical Genome Assembly [22] to yield 1 contiguous DNA sequence for the chromosome and two contiguous DNA sequences for two plasmids.

Bioinformatics analyses

Raw sequence reads were subjected to quality control using PRINSEQ version 0.20.3 [23] with the following metrics: 22 bases from the 5'end and 10 bases from the 3'end were trimmed, in order to reach a minimum mean quality score of 25. Lyve-SET v1.1.4f [24] is a hqSNP pipeline designed to identify SNPs with high quality while removing lower quality SNPs from its analysis, resulting in a high-confidence phylogeny. Reads passing quality control were used as input for the Lyve-SET and *de-novo* genome assembly by SPAdes genome assembler v3.9.0 [25]. Lyve-SET was applied using the reference genome with the following conditions: phages in the reference genome were masked, clustered SNPs within 5 bp were filtered, SNPs calls had to be supported by at least 20x coverage and 95% consensus. Phylogenetic trees inferred by Randomized Axelerated Maximum Likelihood (RAxML) tool [26] integrated in the Lyve-SET pipeline were viewed and formatted in MEGA6 [27]. Phylogenetic trees were midpoint rooted and the taxa were formatted for balanced shape. Lyve-SET also generates a pairwise matrix output file listing the number of pairwise hqSNPs between the genomes in this study. SPAdes was run using the following conditions:—careful (to reduce the number of mismatches and short indels) and—only-assembler (runs assembly module only).

We performed CG-SNP analysis of all the genomes against the reference genome using Parsnp version 1.2, a rapid conservative core genome multi-alignment tool, using default parameters [28]. By default, Parsnp calculates the Maximal Unique Matches Index (MUMi) distances between the reference and each of the genomes in the genome library. MUMi is a genomic distance based on the amount of exact matches at least 19 bp long shared between

two genomes [29]. It only includes genomes with MUMi distance < = 0.01 and excludes genomes that are outside of this parameter. We used the assembled genomes of all ETEC O6 isolates in the genome directory as input for CG alignment by Parsnp. Parsnp yielded a phylogenetic tree that was then viewed and formatted in MEGA6.

We compared the trees resulting from Parsnp and Lyve-SET using a method that we previously developed [24]. Briefly, 10,000 trees with random topologies were created for each comparison, with either the Parsnp tree or the Lyve-SET tree designated as the reference or the query tree. The observed Robinson-Foulds values were calculated, and the average values for each metric were recorded between the random trees and the reference tree. We compared the observed and background distribution with the Z-test to determine how significantly different the observed tree distance was between trees. Next, we approximated the difference in branch length of Clade II in CG-SNP and WG-hqSNP phylogenies by subtraction of the SNPs between the two phylogenies.

We searched the National Center for Biotechnology Information (NCBI) Pathogen Detection Pipeline (https://www.ncbi.nlm.nih.gov/pathogens) for any related genomes to the set of genomes in clade IV (S1 Fig). The Pathogen Detection Pipeline organizes all publicly available genomes on NCBI that are closely related into trees based on the Jaccard distance and these genomes are refined into subclades using SNPs found among related assemblies [24].

Antimicrobial susceptibility testing

Antimicrobial susceptibility testing (AST) was performed according to standard NARMS methodology using broth microdilution on the Sensititre Gram-negative susceptibility panel CMV4AGNF (Thermo Scientific, Inc) [30].

Inoculum was prepared by generating a 0.5 McFarland suspension in 5ml sterile demineralized water for each isolate. A 10μl aliquot of water inoculum was then transferred to 11ml of cation adjusted Mueller Hinton Broth. A 50μl aliquot of broth inoculum suspension was transferred into each well of the 96-well Gram-negative susceptibility panel using the Sensititre AIM automated inoculation delivery system. Panels were incubated at 35°C for 18–20 hours. Results were read using the automated reading and incubation system (ARIS, Thermo Scientific Inc).

The Gram negative panel was comprised of 14 antimicrobial agents and the interpretive criteria for being susceptible or resistant to a drug was followed as previously described [30]. The criteria used to categorize minimum inhibitory concentration (MIC) results as susceptible, intermediate or resistant are based on current guidelines provided by the Clinical Laboratory Standards Institute (CLSI) where available or based on NARMS consensus epidemiologic cutoffs (for streptomycin) [30, 31]. A pansusceptible isolate is defined as an isolate susceptible to the Gram negative panel of 14 antimicrobial agents while a multidrug resistant isolate is defined as resistance to 3 or more CLSI drug classes [31].

Identification of antimicrobial resistance determinants

Assembled genomes of all the isolates in this study (S1 Table) were input into the ResFinder 2.1 tool in the CGE website (https://cge.cbs.dtu.dk) for identification of acquired AR genes, in August 2017. Default thresholds of 90% identity and 60% gene coverage were employed [30, 32]. Additional variants of the *mcr* and *qnr* mobile resistance determinants that confer resistance to polymyxin (e.g. colistin) and fluoroquinolone (e.g. ciprofloxacin) classes of antibiotics [33, 34] were discovered after the initial screening. Determinants present on contigs with low sequencing coverage (defined as less than 5x) were omitted from further analysis as these are sometimes the result of carryover contamination or sample bleed over on the Illumina

platform [35]. To this end, we ran a MegaBLAST using NCBI-blast+ with a 90% identity cutoff to detect *mcr-3*, *mcr-4 and qnrE* determinants in the assembled genomes while the other variants of these genes were screened by ResFinder 2.1.We used an in-house script to extract sequences encoding DNA gyrase (*gyrA*, *gyrB*) and topoisomerase IV (*parC*) (https://github.com/lskatz/ETEC-O6) from the assembled genomes. Multiple sequence alignment in MEGA6 by ClustalW of the translated DNA sequence was used for the detection of mutations with particular emphasis on mutations known to confer resistance in the quinolone resistance determining regions (QRDRs) of these genes [e.g. mutations occurring at $gyrA(S^{83})$, $gyrA(D^{87})$, $parC(S^{80})$] [36–38].

Phenotypic antimicrobial resistance and resistance determinants were visualized alongside WG-hqSNP phylogeny using the ggtree package in R v. 3.5.1 [39].

Identification of virulence factors

Assembled genomes of all the isolates in this study (S1 Table) were input into the Virulence-Finder 1.5 tool in the CGE website (https://cge.cbs.dtu.dk) for identification of enterotoxin virulence genes [40].

Results

Whole genome high-quality SNP phylogeny of ETEC O6 isolates

We analyzed a total of 40 ETEC O6 isolates, 27 of which were from 18 outbreaks and 13 from sporadic cases spanning from 1975–2016 collected from Central America, USA and Guatemala (S1 Table). We employed WG-hqSNP analysis by Lyve-SET to construct the WG phylogeny of the 40 ETEC O6 isolates. The three clade IV genomes were 2,840–7,657 pairwise hqSNPs different from the other genomes included in this study, indicating that these three genomes were an outgroup (S1 Fig) and hence too diverse to be included in the further phylogenetic analyses.

Results from WG-hqSNP analysis indicated that ETEC O6 isolates clustered into three major clades (I, II and III) (Fig 1). Clade I comprised of 9 outbreak and 8 sporadic isolates, clade II comprised of 5 outbreak isolates and clade III comprised of 10 outbreak and 6 sporadic isolates. Clade I isolates differed by 0–840 pairwise hqSNPs with a median of 372 pairwise hqSNPs, clade II isolates differed by 2–215 pairwise hqSNPs with a median of 7 pairwise hqSNPs and clade III isolates differed by 1–471 pairwise hqSNPs with a median of 177 pairwise hqSNPs.

Core genome SNP phylogeny of ETEC O6 isolates

ETEC O6 genomes clustered into three major clades (I, II and III) by CG-SNP analysis performed using Parsnp (Fig 2). Each of the three clades consisted of the same outbreak and sporadic infection isolates as seen in the phylogeny by WG-hqSNP analysis (Fig 1). Clade I genomes differed by 71–1037 pairwise SNPs with a median of 703 pairwise SNPs, clade II genomes differed by 377–542 pairwise SNPs with a median of 423 pairwise SNPs and clade III genomes differed by 160–1007 pairwise SNPs with a median of 601 pairwise SNPs (Fig 2). By default Parsnp excluded the three outgroup genomes (clade IV, S1 Fig) in its analysis because the MUMi distances of these genomes were > 0.01.

Comparison of whole genome high-quality SNP and core genome SNP phylogenies

Overall, the WG-hqSNP and CG SNP phylogenies of ETEC O6 isolates in this study were similar (Figs 1 and 2). When comparing the topologies of both trees, we found that they are more similar than what we would expect by chance alone (Robinson-Foulds metric of 30, $p < 1e$-
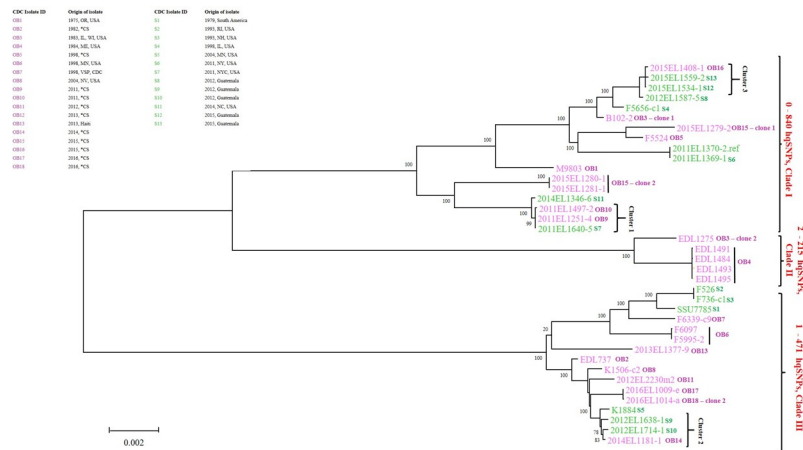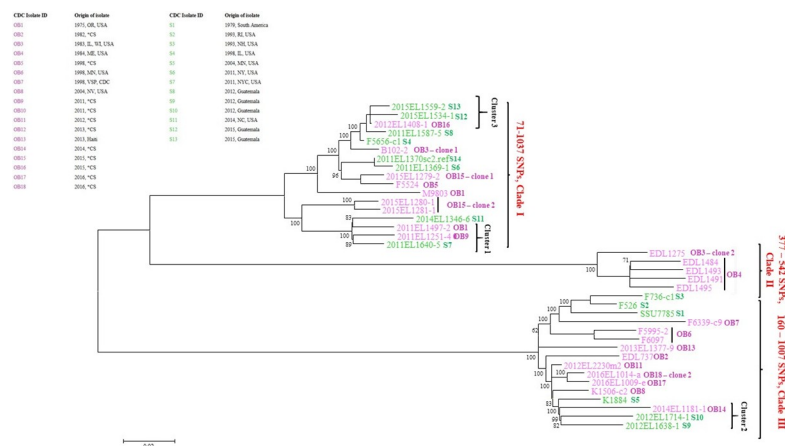
Fig 1. ETEC O6 isolates are clustered into three clades by whole genome high-quality SNP analysis. Whole genome high-quality single nucleotide polymorphisms (WG-hqSNP) tree generated by Lyve-SET for phylogenetic relatedness of 37 ETEC O6 isolates against the reference genome 2011EL1370-2. The scale represents a distance of 0.002 hqSNPs per site. At each ancestor node, bootstrap percentages are displayed. Isolates are clustered into Clade I, Clade II and Clade III. Isolates are color coded based on isolation during an outbreak (OB) or sporadic infection (S). In the metadata table, *CS stands for Cruise Ship; US states are abbreviated.

https://doi.org/10.1371/journal.pone.0208735.g001

99). However, the branch length of clade II in CG-SNP phylogeny was longer when compared to the branch length of clade II in WG-hqSNP phylogeny. We further investigated and attributed the difference in this branch length to 871 pairwise SNPs present in the CG-SNP phylogeny that were not detected in WG-hqSNP phylogeny.

Median number of whole genome high-quality SNPs in ETEC O6 isolates associated with outbreaks

Detecting clusters of isolates from cases that could be associated with a common source using WGS data requires an understanding of within cluster genome diversity for the establishment

Fig 2. ETEC O6 isolates are clustered into three clades by core genome SNP analysis. Core genome single nucleotide polymorphisms (CG-SNP) tree generated by Parsnp for phylogenetic relatedness of 37 ETEC O6 isolates against the reference genome 2011EL1370-2. The scale represents a distance of 0.02 SNPs per site. Bootstrap percentages are displayed at each ancestor node. Isolates are clustered into Clade I, Clade II and Clade III. Isolates are color coded based on isolation during an outbreak (OB) or sporadic infection (S). In the metadata table, *CS stands for Cruise Ship; US states are abbreviated.

https://doi.org/10.1371/journal.pone.0208735.g002

Table 1. Median number of pairwise whole genome high-quality SNPs in ETEC O6 outbreaks are 0–4.5.

Outbreak	Number of isolates	Median number of hqSNPs [1]	Predicted outbreak type
OB3 –clone 1	1	-	Polyclonal
OB3 –clone 2	1	-	Polyclonal
OB4	4	4.5	Clonal
OB6	2	1	Clonal
OB9, OB10	2	4	Clonal
OB12	2	0	Clonal
OB15 –clone 1	1	-	Polyclonal
OB15 –clone 2	2	0	Polyclonal
OB17, OB18-clone 1	2	1	Clonal
OB18 –clone 2	1	-	Polyclonal

OB stands for outbreak. [1] A dash in this column indicates that only one isolate for this clone was available for study and therefore the median could not be calculated.

https://doi.org/10.1371/journal.pone.0208735.t001

of approximate SNP or allele cut-off thresholds. A clonal outbreak is caused by a single strain while a polyclonal outbreak is caused by multiple genetically different strains of the same serotype, multiple serotypes of the same species or multiple species. With the underlying goal of ascertaining clonal or polyclonal ETEC O6 outbreaks in the future based on the pairwise hqSNP counts, we next determined the median number of pairwise hqSNPs in genomes of ETEC O6 isolates belonging to epidemiologically confirmed outbreaks. The median number of pairwise hqSNPs in OB4, OB6, OB9, OB10 and OB12 outbreaks were between 0–5 implying that these were clonal outbreaks (Table 1). In OB15, median number of pairwise hqSNPs in clone 2 was 0 while 759–835 pairwise hqSNPs were detected between isolates belonging to clones 1 and 2 insinuating that OB15 was a polyclonal outbreak. In total, 1,908 pairwise hqSNPS were detected between isolates belonging to clones 1 and 2 of OB3 suggesting that this was also a polyclonal outbreak. Median number of pairwise hqSNPs for OB17 and clone 1 of OB18 was 1 while 5,436 pairwise hqSNPs were identified between clones 1 and 2 of OB18 indicating polyclonality. Overall, based on this study results the median number of pairwise hqSNPs in clonal ETEC O6 outbreaks was less than 5.

Few sporadic infections isolates in clades I and III have clustered closely with outbreak isolates in three clusters where the median number of pairwise hqSNPs were between 4–14 (Table 2).

OB stands for outbreak; S stands for sporadic infection.

Table 2. Median number of pairwise whole genome high-quality SNPs in clusters of isolates from sporadic infections and outbreaks.

Cluster number	CDC isolate ID (sporadic infection/outbreak number)	Median number of pairwise WG-hqSNPs
1	2011EL1640-5 (S7), 2011EL1251-4 (OB 9), 2011EL1497-2 (OB10)	4
2	2012EL1714-1 (S8), 2012EL1638-1 (S8), 2014EL1181-1 (OB14)	14
3	2015EL1534-1 (S10), 2015EL1559-2 (S10), 2015EL1408-1 (OB16)	9

OB stands for outbreak; S stands for sporadic infection.

https://doi.org/10.1371/journal.pone.0208735.t002

Table 3. Phenotypic and genotypic antimicrobial resistance profiles of ETEC O6 isolates.

CDC Isolate ID	Outbreak (OB)/Sporadic (S) infection number	Phenotypic resistance resistance profile	Resistance determinants
F5524	OB5	A,S,Su,T,Cot	bla_{TEM-1B}, strA, strB, sul2, tetA, dfrA8
F5995	OB6	A,S,Su,Cot	bla_{TEM-1B}, strA, strB, sul2, dfrA8
F6097	OB6	A,S,Su,Cot	bla_{TEM-1B}, strA, strB, sul2, dfrA8
2011EL1251-4	OB9	Nal	gyrA83
2011EL1497-2	OB10	Nal	gyrA83
2013EL1319-2	OB12	Nal	gyrA83
2013EL1320-5	OB12	Nal	gyrA83
2013EL1377-9	OB13	S,Su	bla_{TEM-1B}, strA, strB, sul2
2015EL1279-2	OB15	A,Su,Cot	bla_{TEM-1B}, aadA1, sul2, dfrA1, qnrS1
2015EL1280-1	OB15	Nal	bla_{TEM-1B}, gyrA83, strA, strB, sul2, dfrA8
2015EL1281-1	OB15	Nal	gyrA83
2015EL1408-1	OB16	A,S,Su,Cot	bla_{TEM-1B}, strA, strB, sul2, tetB, dfrA8
2016EL1009-e	OB17	Nal	gyrA87
2016EL1012-b	OB18	Nal,Su,T,Cot	gyrA83, aadA1, sul1, tetA, dfrA15
2016EL1014-a	OB18	Nal	gyrA87
SSU7785	S1	S,T	strA, strB, tetB
2011EL1640-5	S7	Nal	gyrA83
2012EL1587-5	S8	A,S,Su,T,Cot	bla_{TEM-1B}, strA, strB, sul2, tetB, dfrA8
2012EL1638-1	S9	Nal	gyrA83
2014EL1346-6	S11	Nal	gyrA83
2015EL1534-1	S12	A,S,Su,T,Cot	bla_{TEM-1B}, strA, strB, sul2, tetB, dfrA8
2015EL1559-2	S13	A,S,Su,T,Cot	bla_{TEM-1B}, strA, strB, sul2, tetB, dfrA8

https://doi.org/10.1371/journal.pone.0208735.t003

Antimicrobial resistance. All isolates in the study were analyzed for antimicrobial susceptibility. Eighteen ETEC O6 isolates were pansusceptible, and the remaining 22 isolates were resistant to one or more antibiotics (Table 3). Heat map visualization of phenotypic resistance and genotypic resistance determinants are shown (Fig 3). Out of the resistant isolates, nine were multidrug resistant. Among the resistant isolates, phenotypic resistance was observed to one or more of the following drugs: ampicillin, nalidixic acid, streptomycin, sulfonamides, tetracycline, trimethoprim-sulfamethoxazole. All isolates were susceptible to amoxicillin-clavulanic acid, azithromycin, cefoxitin, ceftriaxone, chloramphenicol, ciprofloxacin, gentamicin and meropenem. Resistance was correlated in 22 isolates with one or more of the following genotypic AR determinants: bla_{TEM-1B}, aadA1, strA, strB, sul1, sul2, tetA, tetB, dfrA1, dfrA8, dfrA15, and mutations in gyrA. The qnrS1 gene was detected in one isolate lacking phenotypic resistance to quinolones. No known mcr mediated colistin resistance determinants were found.

Discrepancy between phenotypic AR and genotypic AR determinants were detected in 5 ETEC O6 isolates. 2015EL1279-2 was susceptible to streptomycin (MIC = 8 μg/mL) and ciprofloxacin (MIC = 0.25 μg/mL) while aadA1 and qnrS1 were detected; 2015EL1280-1 was susceptible to ampicillin (MIC = 2 μg/mL), streptomycin (MIC = 8 μg/mL), sulfisoxazole (MIC < = 16 μg/mL) and trimethoprim-sulfamethoxazole (MIC < = 0.12/2.38 μg/mL) while bla_{TEM-1B}, strA, strB, sul2 and dfrA8 were detected; 2013EL1377-9 was susceptible to ampicillin (MIC = 2 μg/mL) while bla_{TEM-1B} was detected; 2012EL1408-1 was susceptible to tetracycline while tetB (MIC < = 4 μg/mL) was detected and 2016EL1012-b was susceptible to streptomycin (MIC = 16 μg/mL), yet an aadA1 gene was present in this isolate (Table 3).
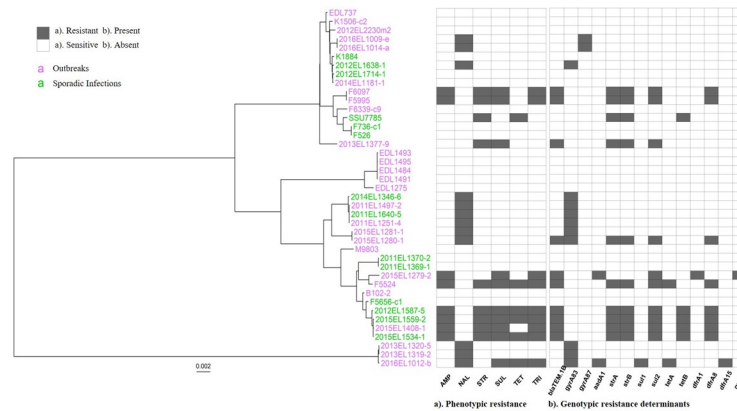
Fig 3. Heat map visualization of phenotypic and genotypic antimicrobial resistance patterns of ETEC O6 isolates.
Phenotypic antimicrobial resistance and resistance determinants were visualized alongside WG-hqSNP phylogeny
using the ggtree package in R. a). Phenotypic antimicrobial resistance patterns: AMP: Ampicillin; NAL: Nalidixic acid;
STR: Streptomycin; SUL: Sulfisoxazole; TET: Tetracycline; TRI: trimethoprim-sulfamethoxazole. b). Genotypic
resistance determinants: bla_{TEM-1B}: beta-lactam resistance determinant; $gyrA83$, $gyrA87$: point mutations in S83L and
D87Y in Gyrase A subunit of DNA Gyrase enzyme resulting in reduced binding of nalidixic acid; $aadA1$, $strA$, $strB$:
streptomycin resistant determinants; $sul1$, $sul2$: sulphonamide resistance determinant; $tetA$, $tetB$: tetracycline resistance
determinant; $dfrA1$, $dfrA8$ and $dfrA15$: trimethoprim-sulfamethoxazole resistance determinant; $qnrS1$: quinolone
resistance determinant.

https://doi.org/10.1371/journal.pone.0208735.g003

Phenotypic resistance

A: Ampicillin; Nal: Nalidixic acid; S: Streptomycin; Su: Sulfisoxazole; T: Tetracycline; Cot: tri-
methoprim-sulfamethoxazole. bla_{TEM-1B}: beta-lactam resistance determinant; $gyrA83$, $gyrA87$:
point mutations in S83L and D87Y in Gyrase A subunit of DNA Gyrase enzyme resulting in
reduced binding of nalidixic acid; $aadA1$, $strA$, $strB$: streptomycin resistant determinants; $sul1$,
$sul2$: sulphonamide resistance determinant; $tetA$, $tetB$: tetracycline resistance determinant;
$dfrA1$, $dfrA8$ and $dfrA15$: trimethoprim-sulfamethoxazole resistance determinant; $qnrS1$: quin-
olone resistance determinant.

All isolates were susceptible to CLA: Amoxicillin-clavulanic acid, AZT: Azithromycin, CEF:
Cefoxitin, CEFT: Ceftriaxone, CHL: Chloramphenicol, CIP: Ciprofloxacin, GEN: Gentamicin
and MER: Meropenem.

The following isolates were pansusceptible to all antibiotics tested by antimicrobial suscep-
tibility testing and its genomes had no antibiotic resistance determinants: K1884, F6339-c9,
2012EL2230m2, B102-2, 2014EL1181-1, 2011EL1369-1, 2012EL1714-1, F526, F736-c1,
EDL1493, EDL1495, EDL1484, EDL1491, F5656-c1, EDL737, EDL1275, K1506-c2, M9803,
2011EL1370-2 (reference).

Virulence factors

Thirty-two isolates were predicted to harbor heat-labile enterotoxin A subunit *(ltcA)* and heat-
stable enterotoxin 1 *(astA)* virulence factors; 6 and 2 isolates were predicted to harbor *astA*
only and *ltcA* only respectively (S1 Table).

Discussion

Globally, O6 is the most common ETEC O serogroup [8, 14]; it is therefore important to char-
acterize O6 strains phylogenetically to understand the evolution and genetic diversity. In this
study, we performed genome wide characterization of historical and recently collected ETEC
O6 isolates in order to understand the phylogenetics of this serogroup, which has a significant

public health impact. Effective subtyping by WGS during future outbreaks will pave the way to quicker outbreak investigations that encompasses creation of an epidemic curve, case definition, and detection of the infection source ultimately resulting in the containment of the outbreak. This study is unique because, to our knowledge, it is the first to compare whole genome and core genome SNP analyses, determine the median number of pairwise hqSNPs in a clonal outbreak and characterize antimicrobial resistance of ETEC O6 isolates. Information gathered in this study will be part of a larger data set to be used to validate whole genome multi-locus sequence typing (wgMLST) for *E. coli* by PulseNet USA and PulseNet International [41].

Our results indicate that ETEC O6 isolates are genetically diverse, grouped into 3 clades by WG-hqSNP and CG SNP analyses independent of the time of isolation. WG-hqSNP identified the three outgroup genomes (clade IV, S1 Fig) that were removed from downstream WG-hqSNP and automatically excluded from CG-SNP analyses. These genomes were also diverse from other publicly available *E. coli* and *Shigella* genomes. Furthermore, these were confirmed as ETEC O6 by sequence based methods. As *E. coli* is subjected to homologous and non-homologous recombination [42], we speculate recombination to be the reason for extensive genetic diversity seen in the aforementioned genomes while selective pressure to maintain the lipopolysaccharide (LPS) component of the outer membrane (OM) of these isolates could have resulted in preservation of the O6 antigen. Phylogenetic relatedness between the O6 isolates is similar by WG-hqSNP and CG-SNP analyses indicating that these are likely to be true relationships (Figs 1 and 2). ETEC has been shown to partially cluster by phylogenetic analysis based on the O-antigen [14]. Therefore, the robustness of the phylogenetic analyses of ETEC O6 isolates by WG-hqSNP and CG-SNP analyses seen in this study will most likely not be evident if isolates of different O-serogroups were included.

The number of pairwise SNPs is consistently higher by CG-SNP analysis (Fig 2) when compared to WG-hqSNP analysis (Fig 1) in all three clades possibly because Lyve-SET is very conservative when calling nucleotide variants and as a result, the percentage of called variants reported is usually lower than those from Parsnp. Additionally, unlike Lyve-SET, which uses pre-processed raw sequencing reads as input, Parsnp uses assembled genomes where there are opportunities for errors to occur during assembly that could contribute to the higher SNP count.

Our results show that median number of pairwise hqSNPs in a clonal ETEC O6 outbreak is less than 5 (Table 1). We observed three clusters where isolates from sporadic infections closely clustered with outbreak isolates (Table 2 and Fig 1). Isolates within the first and third clusters are temporally associated so a possible common source cannot be excluded. The other possibility is the existence of common widely distributed sequence types. While the two isolates from Guatemala in the second cluster are geographically and temporally associated, the outbreak related isolate in that cluster was recovered two years later so this case is more likely to represent a common clone. Our study findings seem to be in agreement with the conclusions of von Mentzer *et al* [14] and Sahl *et al* [15] where genetic similarity of ETEC clones were identified in strains isolated from temporally and geographically dispersed cases of ETEC mediated diarrhea. The number of samples is small due to lack of routine ETEC surveillance in the US and the depth of sampling for each outbreak is also small. With a larger sample size, we would expect the within outbreak variability to increase slightly.

Three of the outbreaks were deemed to be polyclonal in nature, i.e., caused by more than one strain of ETEC (Table 1, Figs 1 and 2). The isolates from OB3 were from neighboring states and both cases reported consuming the same cheese product from a local creamery (S1 Table). OB15 and OB18 both took place on cruise ships. The majority of the cruise ship outbreaks investigated in the USA are polyclonal. Typically, the source is a food that is collected from one of the harbors the ship visits to replenish vessel's food supply. Cruise ships of the

outbreaks in this study have mostly sailed on the Caribbean Sea and/or South America. Countries along this cruise route do not have as rigorous food safety standards as the USA, therefore foods can be heavily contaminated with multiple strains of the same species/serotype or even multiple species, which is the definition of a polyclonal outbreak [43]. We also attribute polyclonality as the reason for different resistance profiles of clone 1 and clone 2 isolates of OB15 (Fig 3).

Twenty-two ETEC O6 isolates were resistant to one or more antibiotics out of which 9 were multi-drug resistant (Table 2). This is in line with results from Medina *et al* [11] where ETEC from Peruvian children were also resistant to older inexpensive antibiotics such as ampicillin, trimethoprim-sulfamethoxazole and tetracycline while being susceptible to ciprofloxacin and cephalosporins. Since many AR determinants are on plasmids or other mobile elements [14], there are opportunities for the loss of mobile elements during sub-culturing for testing. In this study, WGS was conducted prior to AST and there were chances for AR determinants to be lost by sub-culturing between sequencing and AST. We attribute this to be the reason for some discrepancies observed between phenotypic AR and resistant determinants in this study. 2015EL1280-1 was phenotypically susceptible to ampicillin (MIC = 2 μg/ml), streptomycin (MIC = 8 μg/ml), sulfisoxazole (MIC $<$ = 16 μg/ml) and trimethoprim-sulfamethoxazole (MIC $<$ = 0.12/2.38 μg/ml) while the corresponding encoding elements were present in these genomes. The resistance determinants were present at 100% nucleotide identity and 100% gene coverage when compared to their best match in ResFinder, but were on a contig with sequencing coverage that was approximately 25% of the average genome coverage (contig coverage 19x; average genome coverage 80x). This observation could be the result of the loss of these resistance determinants in some cells that underwent sequencing and subsequent susceptibility testing. Similarly, 2012EL1408-1 was susceptible to tetracycline (MIC was $<$ = 4 μg/ml) while *tetB* was present in its genome. The *tetB* gene was nearly identical to the reference in ResFinder, with a single amino acid change (P$>$H) present at amino acid 390 (data not presented). 2013EL1377-9 possessed a *bla*TEM-1B gene in the absence of ampicillin resistance, however upon further examination it was observed that only 63% percent of the gene was present due to an interruption by IS*26* (data not presented), which likely explains the phenotype observed. 2016EL1012-b was sensitive to streptomycin, yet an *aadA1* gene was present in this isolate at 96.46% identity and 100% gene coverage. Closer examination did not reveal any premature stop codons, but did reveal several non-synonymous mutations in the coding sequence (data not presented). The MIC to streptomycin was 16 μg/mL. The US National Antimicrobial Resistance Monitoring System (NARMS) employs an epidemiological cutoff value (ECV) of $\geq$ 32 μg/mL for streptomycin as there are no standardized breakpoints for Enterobacteriaceae. While most isolates with streptomycin MICs below this ECV do not possess resistance determinants, *E. coli* isolates with an *aadA* gene have been encountered with MICs of 8 and 16 μg/mL [44].

Finally, *aad*A1 and *qnrS1* were detected in 2015EL-1279-2 at 100% nucleotide identity and gene coverage in the absence of streptomycin or quinolone resistance (Table 2). As noted previously, while most isolates with streptomycin MICs below the ECV for this drug do not possess resistance determinants, *E. coli* isolates with an *aadA* gene have been encountered with lower MIC values [44]. Plasmid-mediated quinolone resistance genes (PMQR) such as the *qnr* gene has been observed to confer low level fluoroquinolone resistance (i.e. ciprofloxacin), with little impact on nalidixic acid MICs[31]. PMQR genes in association with chromosomal resistance mechanisms significantly increase the resistance to quinolones and fluoroquinolones. The decreased susceptibility conferred by *qnr* genes is generally not sufficient to raise the MIC over the current CLSI upper limit of the susceptible range ($\leq$ 1 μg/mL for *E. coli* and *Shigella.*) [34]. The MIC for ciprofloxacin in 2015EL-1279-2 was 0.25 μg/mL which is in the expected

range for *qnr+ E. coli* [45] and this genome did not harbor chromosomal mutations to *gyrA* and/or *parC* genes in the QRDR.

In conclusion, ETEC O6 isolates in this study showed significant genomic diversity [46] by phylogenetic analyses independent of time and geographical location of isolation. However, a few clonal subtypes causative of an outbreak and/or sporadic infections were observed. The phylogenetic relationships between the isolates were similar by WG-hqSNP and CG SNP analyses thereby providing support that they represent true evolutionary relationships. Based on our results, fewer than 5 pairwise hqSNPs between ETEC O6 isolates is circumstantial evidence of an outbreak cluster. This study provides a baseline for understanding the ETEC O6 population structure by WG-hqSNP and CG-SNP analyses as PulseNet USA moves forward with implementation of routine surveillance for ETEC using WGS. Our study identified multi-drug resistance among 6 of 27 outbreak isolates and 3 of 13 sporadic isolates, a finding that is concerning and warrants further investigation. Similar genome wide characterization studies of other dominant circulating serogroups of ETEC and/or other non Shiga toxin producing *E. coli* such as enteropathogenic *E. coli* (EPEC), enteroaggregative *E. coli* (EAEC) and diffuse-adhering enteroinvasive *E. coli* (EIEC) are needed to understand their population structure and AR patterns.

Supporting information

S1 Fig. ETEC O6 isolates are clustered into four clades by WG-hqSNP analysis. Whole genome high-quality single nucleotide polymorphisms (WG-hqSNP) analysis tree generated by Lyve-SET for phylogenetic relatedness of 40 ETEC O6 isolates against the reference genome 2011EL1370-2. The scale represents a distance of 0.002 hqSNPs per site. At each ancestor node, bootstrap percentages are displayed. Isolates clustered into 4 major clades. Isolates are clustered into Clade I, Clade II and Clade III. Isolates are color coded based on isolation during an outbreak (OB) or sporadic infection (S). In the metadata table, *CS stands for Cruise Ship; US states are abbreviated.
(TIF)

S1 Table. Bacterial isolates used in this study.
(DOCX)

Author Contributions

Conceptualization: Vaishnavi Pattabiraman, Lee S. Katz, Jessica C. Chen, Eija Trees.

Formal analysis: Vaishnavi Pattabiraman, Lee S. Katz, Jessica C. Chen.

Investigation: Vaishnavi Pattabiraman, Lee S. Katz, Jessica C. Chen, Eija Trees.

Methodology: Vaishnavi Pattabiraman, Lee S. Katz, Jessica C. Chen, Andre E. McCullough, Eija Trees.

Supervision: Eija Trees.

Writing – original draft: Vaishnavi Pattabiraman.

Writing – review & editing: Lee S. Katz, Jessica C. Chen, Eija Trees.

References

1. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med. 2015; 12(12):e1001921. https://doi.org/10.1371/journal.pmed.1001921 PMID: 26633831; PubMed Central PMCID: PMCPMC4668831.

2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. 2013; 382(9888):209–22. https://doi.org/10.1016/S0140-6736(13)60844-2 PMID: 23680352.

3. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE, et al. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. PLoS One. 2013; 8(9): e72788. https://doi.org/10.1371/journal.pone.0072788 PMID: 24023773; PubMed Central PMCID: PMCPMC3762858.

4. Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt H, Steinsland H. Molecular mechanisms of enterotoxigenic Escherichia coli infection. Microbes Infect. 2010; 12(2):89–98. https://doi.org/10.1016/j.micinf.2009.10.002 PMID: 19883790.

5. Gaastra W, Svennerholm AM. Colonization factors of human enterotoxigenic Escherichia coli (ETEC). Trends Microbiol. 1996; 4(11):444–52. PMID: 8950814.

6. Wolf MK. Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic Escherichia coli. Clin Microbiol Rev. 1997; 10(4):569–84. PMID: 9336662; PubMed Central PMCID: PMCPMC172934.

7. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev. 2005; 18(3):465–83. https://doi.org/10.1128/CMR.18.3.465-483.2005 PMID: 16020685; PubMed Central PMCID: PMCPMC1195967.

8. Shin J, Yoon KB, Jeon DY, Oh SS, Oh KH, Chung GT, et al. Consecutive Outbreaks of Enterotoxigenic Escherichia coli O6 in Schools in South Korea Caused by Contamination of Fermented Vegetable Kimchi. Foodborne Pathog Dis. 2016; 13(10):535–43. https://doi.org/10.1089/fpd.2016.2147 PMID: 27557346.

9. Pattabiraman V, Bopp CA. Draft Whole-Genome Sequences of 10 Serogroup O6 Enterotoxigenic Escherichia coli Strains. Genome Announc. 2014; 2(6). Epub 2014/12/20. https://doi.org/10.1128/genomeA.01274-14 PMID: 25523765; PubMed Central PMCID: PMCPMC4271155.

10. Kwon T, Chung SY, Jung YH, Jung SJ, Roh SG, Park JS, et al. Comparative genomic analysis and characteristics of NCCP15740, the major type of enterotoxigenic Escherichia coli in Korea. Gut Pathog. 2017; 9:55. https://doi.org/10.1186/s13099-017-0204-y PMID: 28943892; PubMed Central PMCID: PMCPMC5607484.

11. Medina AM, Rivera FP, Pons MJ, Riveros M, Gomes C, Bernal M, et al. Comparative analysis of antimicrobial resistance in enterotoxigenic Escherichia coli isolates from two paediatric cohort studies in Lima, Peru. Trans R Soc Trop Med Hyg. 2015; 109(8):493–502. https://doi.org/10.1093/trstmh/trv054 PMID: 26175267; PubMed Central PMCID: PMCPMC4592336.

12. Begum YA, Talukder KA, Azmi IJ, Shahnaij M, Sheikh A, Sharmin S, et al. Resistance Pattern and Molecular Characterization of Enterotoxigenic Escherichia coli (ETEC) Strains Isolated in Bangladesh. PLoS One. 2016; 11(7):e0157415. https://doi.org/10.1371/journal.pone.0157415 PMID: 27428376; PubMed Central PMCID: PMCPMC4948870.

13. CDC. Enterotoxigenic E. coli 2014. Available from: www.cdc.gov/ecoli/etec.html.

14. von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, et al. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. Nat Genet. 2014; 46 (12):1321–6. https://doi.org/10.1038/ng.3145 PMID: 25383970.

15. Sahl JW, Sistrunk JR, Fraser CM, Hine E, Baby N, Begum Y, et al. Examination of the Enterotoxigenic Escherichia coli Population Structure during Human Infection. MBio. 2015; 6(3):e00501. https://doi.org/10.1128/mBio.00501-15 PMID: 26060273; PubMed Central PMCID: PMCPMC4462620.

16. Mehla K, Ramana J. Identification of epitope-based peptide vaccine candidates against enterotoxigenic Escherichia coli: a comparative genomics and immunoinformatics approach. Mol Biosyst. 2016; 12 (3):890–901. https://doi.org/10.1039/c5mb00745c PMID: 26766131.

17. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. J Clin Microbiol. 2015; 53

(8):2410–26. Epub 2015/05/15. https://doi.org/10.1128/JCM.00008-15 PMID: 25972421; PubMed Central PMCID: PMCPMC4508402.

18. WH E. The genus Escherichia. In: Edwards and Ewing's identification of Enterobacteriaceae, 4th ed. Elsevier Science Publishing Co. 1986:93–134.

19. CDC. Laboratory standard operating procedure for PulseNet Nextera XT library prep and run setup for Illumina Miseq https://www.cdc.gov/pulsenet/pdf/PNL32-MiSeq-Nextera-XT.pdf2016. Available from: https://www.cdc.gov/pulsenet/pdf/PNL32-MiSeq-Nextera-XT.pdf.

20. CDC. Pulsenet standard operating procedue for Illumina MiSeq data quality control [Standard operating procedure (SOP)]. https://www.cdc.gov/pulsenet/pdf/PNQ07_Illumina-MiSeq-Data-QC-508-v1.pdf2015. Available from: https://www.cdc.gov/pulsenet/pdf/PNQ07_Illumina-MiSeq-Data-QC-508-v1.pdf.

21. Smith P, Lindsey RL, Rowe LA, Batra D, Stripling D, Garcia-Toledo L, et al. High-Quality Whole-Genome Sequences for 21 Enterotoxigenic Escherichia coli Strains Generated with PacBio Sequencing. Genome Announc. 2018; 6(2). Epub 2018/01/13. https://doi.org/10.1128/genomeA.01311-17 PMID: 29326203; PubMed Central PMCID: PMCPMC5764927.

22. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013; 10(6):563–9. Epub 2013/05/07. https://doi.org/10.1038/nmeth.2474 PMID: 23644548.

23. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27(6):863–4. https://doi.org/10.1093/bioinformatics/btr026 PMID: 21278185; PubMed Central PMCID: PMCPMC3051327.

24. Katz LS, Griswold T, Williams-Newkirk AJ, Wagner D, Petkau A, Sieffert C, et al. A Comparative Analysis of the Lyve-SET Phylogenomics Pipeline for Genomic Epidemiology of Foodborne Pathogens. Front Microbiol. 2017; 8:375. https://doi.org/10.3389/fmicb.2017.00375 PMID: 28348549; PubMed Central PMCID: PMCPMC5346554.

25. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol. 2013; 20(10):714–37. https://doi.org/10.1089/cmb.2013.0084 PMID: 24093227; PubMed Central PMCID: PMCPMC3791033.

26. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.

27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30(12):2725–9. https://doi.org/10.1093/molbev/mst197 PMID: 24132122; PubMed Central PMCID: PMCPMC3840312.

28. 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(11):524. https://doi.org/10.1186/s13059-014-0524-x PMID: 25410596; PubMed Central PMCID: PMCPMC4262987.

29. Deloger M, El Karoui M, Petit MA. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. J Bacteriol. 2009; 191(1):91–9. Epub 2008/11/04. https://doi.org/10.1128/JB.01202-08 PMID: 18978054; PubMed Central PMCID: PMCPMC2612450.

30. CDC. National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human Isolates Surveillance Report for 2015 (Final Report) https://www.cdc.gov/narms/pdf/2015-NARMS-Annual-Report-cleared_508.pdf: 2018.

31. Institute CaLS. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Eighth Informational Supplement.2018.

32. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012; 67(11):2640–4. Epub 2012/07/12. https://doi.org/10.1093/jac/dks261 PMID: 22782487; PubMed Central PMCID: PMCPMC3468078.

33. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis. 2016; 16(2):161–8. https://doi.org/10.1016/S1473-3099(15)00424-7 PMID: 26603172.

34. Ferreira JC, Filho R, Kuaye APY, Andrade LN, Junior AB, Darini ALC. Identification and characterization of plasmid-mediated quinolone resistance determinants in Enterobacteriaceae isolated from healthy poultry in Brazil. Infect Genet Evol. 2018. https://doi.org/10.1016/j.meegid.2018.02.003 PMID: 29427764.

35. Mitra A, Skrzypczak M, Ginalski K, Rowicka M. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform. PLoS One. 2015; 10(4): e0120520. Epub 2015/04/11. https://doi.org/10.1371/journal.pone.0120520 PMID: 25860802; PubMed Central PMCID: PMCPMC4393298.

36. Reece RJ, Maxwell A. DNA gyrase: structure and function. Crit Rev Biochem Mol Biol. 1991; 26(3–4):335–75. https://doi.org/10.3109/10409239109114072 PMID: 1657531.

37. Peng H, Marians KJ. Escherichia coli topoisomerase IV. Purification, characterization, subunit structure, and subunit interactions. J Biol Chem. 1993; 268(32):24481–90. PMID: 8227000.

38. Drlica K, Zhao X. DNA gyrase, topoisomerase IV, and the 4-quinolones. Microbiol Mol Biol Rev. 1997; 61(3):377–92. PMID: 9293187; PubMed Central PMCID: PMCPMC232616.

39. Yu DS G, Zhu H, Guan Y, Lam TTY. ggtree: a phylogenetic tree viewer for different types of tree annotations. Methods in Ecology and Evolution. 2017; 8(1):28–36.

40. 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(5):1501–10. Epub 2014/02/28. https://doi.org/10.1128/JCM.03617-13 PMID: 24574290; PubMed Central PMCID: PMCPMC3993690.

41. Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Euro Surveill. 2017; 22(23). https://doi.org/10.2807/1560-7917.ES.2017.22.23.30544 PMID: 28662764; PubMed Central PMCID: PMCPMC5479977.

42. Didelot X, Meric G, Falush D, Darling AE. Impact of homologous and non-homologous recombination in the genomic evolution of Escherichia coli. BMC Genomics. 2012; 13:256. Epub 2012/06/21. https://doi.org/10.1186/1471-2164-13-256 PMID: 22712577; PubMed Central PMCID: PMCPMC3505186.

43. Lawrence DN. Outbreaks of Gastrointestinal Diseases on Cruise Ships: Lessons from Three Decades of Progress. Curr Infect Dis Rep. 2004; 6(2):115–23. Epub 2004/03/17. PMID: 15023273.

44. Tyson GH, Li C, Ayers S, McDermott PF, Zhao S. Using whole-genome sequencing to determine appropriate streptomycin epidemiological cutoffs for Salmonella and Escherichia coli. FEMS Microbiol Lett. 2016; 363(4). https://doi.org/10.1093/femsle/fnw009 PMID: 26781915.

45. Hooper DC, Jacoby GA. Mechanisms of drug resistance: quinolone resistance. Ann N Y Acad Sci. 2015; 1354:12–31. https://doi.org/10.1111/nyas.12830 PMID: 26190223; PubMed Central PMCID: PMCPMC4626314.

46. Holmes A, Allison L, Ward M, Dallman TJ, Clark R, Fawkes A, et al. Utility of Whole-Genome Sequencing of Escherichia coli O157 for Outbreak Detection and Epidemiological Surveillance. J Clin Microbiol. 2015; 53(11):3565–73. Epub 2015/09/12. https://doi.org/10.1128/JCM.01066-15 PMID: 26354815; PubMed Central PMCID: PMCPMC4609728.