Genomic Epidemiology and Antimicrobial Susceptibility Profile of Enterotoxigenic *Escherichia coli* From Outpatients With Diarrhea in Shenzhen, China, 2015–2020

Chao Yang[1,2,3†], Yinghui Li[1†], Le Zuo[1†], Min Jiang[1], Xianglilan Zhang[3], Li Xie[4], Miaomiao Luo[5], Yiying She[5], Lei Wang[1], Yixiang Jiang[1], Shuang Wu[1], Rui Cai[1], Xiaolu Shi[1], Yujun Cui[3*], Chengsong Wan[2*] and Qinghua Hu[1*]

[1] Shenzhen Center for Disease Control and Prevention, Shenzhen, China, [2] Department of Microbiology, School of Public Health, Southern Medical University, Guangzhou, China, [3] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China, [4] School of Public Health, University of South China, Hengyang, China, [5] School of Public Health, Shanxi Medical University, Taiyuan, China

Enterotoxigenic *Escherichia coli* (ETEC) is the leading cause of severe diarrhea in children and the most common cause of diarrhea in travelers. However, most ETEC infections in Shenzhen, China were from indigenous adults. In this study, we characterized 106 ETEC isolates from indigenous outpatients with diarrhea (77% were adults aged >20 years) in Shenzhen between 2015 and 2020 by whole-genome sequencing and antimicrobial susceptibility testing. Shenzhen ETEC isolates showed a remarkable high diversity, which belonged to four *E. coli* phylogroups (A: 71%, B1: 13%, E: 10%, and D: 6%) and 15 ETEC lineages, with L11 (25%, O159:H34/O159:H43, ST218/ST3153), novel L2/4 (21%, O6:H16, ST48), and L4 (15%, O25:H16, ST1491) being major lineages. Heat-stable toxin (ST) was most prevalent (76%, STh: 60% STp: 16%), followed by heat-labile toxin (LT, 17%) and ST + LT (7%). One or multiple colonization factors (CFs) were identified in 68 (64%) isolates, with the common CFs being CS21 (48%) and CS6 (34%). Antimicrobial resistance mutation/gene profiles of genomes were concordant with the phenotype testing results of 52 representative isolates, which revealed high resistance rate to nalidixic acid (71%), ampicillin (69%), and ampicillin/sulbactam (46%), and demonstrated that the novel L2/4 was a multidrug-resistant lineage. This study provides novel insight into the genomic epidemiology and antimicrobial susceptibility profile of ETEC infections in indigenous adults for the first time, which further improves our understanding on ETEC epidemiology and has implications for the development of vaccine and future surveillance and prevention of ETEC infections.

Keywords: enterotoxigenic *Escherichia coli* (ETEC), molecular epidemiology, whole genome sequencing (WGS), virulence factor, antimicrobial resistance (AMR), pathogenicity, public health surveillance

INTRODUCTION

Enterotoxigenic *Escherichia coli* (ETEC) is the leading cause of severe diarrhea particularly among young children aged less than five in developing countries, and is also the most common cause of diarrhea in travelers to ETEC-endemic areas, accounting for more than 200 million diarrheal cases and 50,000 deaths annually (Khalil et al., 2018; Fleckenstein and Kuhlmann, 2019). ETEC is defined by production of heat-stable toxin (ST) and/or heat-labile toxin (LT), and ST includes two subtypes, human ST (STh) and porcine ST (STp). STh is the most prevalent enterotoxin associated with human diarrhea, while STp is originally isolated from a porcine source and more prevalent in isolates from animals (So et al., 1976, 1978). There is a remarkable genetic diversity of ST and LT, and multiple variants have been identified (Joffré et al., 2015, 2016). In addition to enterotoxins, most ETEC isolates express one or more plasmid-encoded colonization factors (CF), which are fimbrial or afimbrial surface structures that enable adherence to intestinal epithelium. At least 27 known or putative CFs have been identified to date, including a novel CF, CS30 identified by genomic analysis, and the most prevalent CFs are colonization factor antigen I (CFA/I) and *coli* surface antigens 1–6 (CS1–CS6) (Gaastra and Svennerholm, 1996; Nada et al., 2011; von Mentzer et al., 2017; Vidal et al., 2019). Besides the classic virulence factors of enterotoxins and CFs, multiple non-classic virulence factors have been identified in recent years, including a cytoplasmic protein (LeoA), three adhesins (Tia, TibA, and EtpA), an extracytoplasmic protein (CexE), a hemolysin (ClyA), two mucinases (EatA and YghJ), three iron acquisition systems (Irp1, Irp2, and FyuA), and an enteroaggregative heat-stable toxin 1 (EAST1) (Turner et al., 2006; Pilonieta et al., 2007; Del Canto et al., 2011; Gonzales et al., 2013; Luo et al., 2014; Sjöling et al., 2015).

Enterotoxigenic *Escherichia coli* is genetically highly diverse, and more than 100 serotypes have been identified in clinical isolates (Wolf, 1997; Isidean et al., 2011). Multi-locus sequence type (MLST)-based studies showed that ETEC isolates can be found across five *E. coli* phylogroups including A, B1, B2, D, and E, with A and B1 being more prevalent (Turner et al., 2006; Sahl et al., 2011). Whole genome sequencing (WGS) of a large collection of representative global ETEC isolates further identified 21 robust lineages (L1–L21) characterized by distinct enterotoxin and CF profiles, which belonged to phylogroups A (12 lineages), B1 (7 lineages), C (1 lineage), and E (1 lineage) (von Mentzer et al., 2014; Denamur et al., 2021). Despite the diversity, a clear association between lineage and enterotoxin, CF, serotype, and plasmid content was identified. For example, close related lineages L1 and L2 encoded CS1 + CS3 and CS2 + CS3 (with/without CS21), respectively, but shared common enterotoxin profiles (STh + LT) and O antigen (O6) (von Mentzer et al., 2014). In addition, WGS has been used to characterize the epidemiology of ETEC infections in Bangladesh and Chile, and identified remarkable diversity of local circulating phylogroups and lineages. However, the dominant pathogenic phylogroups were distinct, with more Bangladeshi isolates belonging to phylogroup B1 while most Chilean isolates were from phylogroup A (Sahl et al., 2017; Rasko et al., 2019).

With the widespread use of antimicrobial agents, antimicrobial resistance (AMR) has emerged in ETEC isolates from both children and travelers with diarrhea (Qadri et al., 2005). AMR to commonly used agents such as nalidixic acid (NAL), ampicillin (AMP), tetracycline (TET), and sulfonamides has been frequently detected in ETEC isolates in Peru (Rivera et al., 2010), Bangladesh (Begum et al., 2016), South Korea (Oh et al., 2014), and China (Li et al., 2017), and the emergence of extended-spectrum β-lactamase (ESBL)-producing ETEC poses a new challenge to clinical treatment and public health (Margulieux et al., 2018; Guiral et al., 2019). Moreover, high-level and multidrug-resistant (MDR) had developed in ETEC isolates, which might be related to heavy clinical use of antimicrobial agents (Begum et al., 2016; Li et al., 2017). Due to the increased AMR in many areas over time, azithromycin and fluoroquinolones have been used as the first-line drugs for ETEC infections (Xiang et al., 2020). However, azithromycin-resistant ETEC also emerged in multiple countries (Abraham et al., 2014; Begum et al., 2016; Xiang et al., 2020), and is highly prevalent in Shanghai, China (Xiang et al., 2020), highlighting the necessity of ongoing surveillance of AMR, especially to first-line drugs.

While ETEC mainly causes diarrhea among children and travelers, most ETEC infections in Shenzhen, a populous developed city in southern China, were from indigenous adults, and the genomic epidemiology of Shenzhen ETEC isolates remains unclear. In this study, we sequenced the whole genomes of 106 ETEC isolates from indigenous outpatients with diarrhea in Shenzhen between 2015 and 2020, and compared them to a global collection of representative *E. coli* and ETEC genomes (von Mentzer et al., 2014; Horesh et al., 2021), to characterize the genomic diversity and virulence factors of ETEC in Shenzhen. Moreover, we integrated WGS-based *in silico* AMR mutation/gene detection and antimicrobial susceptibility testing to characterize the AMR profiles of Shenzhen ETEC isolates.

MATERIALS AND METHODS

Strain Collection

ETEC strains were isolated from stool samples of outpatients with diarrhea in 16 sentinel hospitals in Shenzhen, China during routine foodborne disease surveillance between 2015 and 2020 as previously described (Li et al., 2017). Stool samples were enriched or inoculated on selective medium to isolate common foodborne pathogens, including *Salmonella*, *Shigella*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Staphylococcus aureus*, *E. coli* O157:H7, ETEC, enteropathogenic *E. coli*, enteroinvasive *E. coli*, enterohemorrhagic *E. coli*, *Bacillus cereus*, group A *Streptococcus*, and *Listeria monocytogenes*. For ETEC, stool samples were further inoculated on CHROMagar ECC plates and incubated at 37°C overnight. Then, three colonies were randomly selected and inoculated on triple sugar iron agar for incubation and identification by screening ST and LT genes using a modified molecular beacon-based multiplex real-time PCR assay (Chen Q. et al., 2014). ETEC isolates were defined by either ST or LT gene was positive, and a total of 106 ETEC isolates were included in this study.

Whole Genome Sequencing and Genome Dataset

Genomic DNA was extracted using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) according to manufacturer's instructions. Pair-end libraries with a mean insert size of 350 bp were prepared for sequencing using Illumina NovaSeq 6000 platforms. The average read length is 150 bp, and ~1.2 Gb clean data were generated for each isolate on average. Short-read sequencing data of Shenzhen isolates have been deposited in the NCBI Sequence Read Archive under the BioProject PRJNA739477, and the accession numbers were listed in **Supplementary Table 1**.

A total of 177 genomes were analyzed in this study, including 106 newly sequenced genomes of Shenzhen isolates and 71 publicly available representative genomes. These representative genomes were from the largest 50 *E. coli* lineages (Horesh et al., 2021) and 21 global ETEC lineages (von Mentzer et al., 2014), representing the phylogenetic diversity of *E. coli* and ETEC.

Genome Assembly, Annotation, and Pan-Genome Analysis

We performed *de novo* assembling of genomes using SPAdes v3.14.1 (Bankevich et al., 2012) and obtained the assembled genomes of 106 ETEC isolates. The average number of contigs and size of assemblies were 123 (61–244, >500 bp) and 5.0 Mb (4.7–5.2), with an average of 102-fold (88–128) depth for each genome.

The assembled genomes were annotated using Prokka (Seemann, 2014) with default settings, and the gene annotation results (GFF3 files) were used in Panaroo (Tonkin-Hill et al., 2020) to identify the pan-genome and generate the matrix of accessory gene presence/absence. Genes presented in ≥99% isolates were defined as core genes, and the other genes were defined as accessory genes. EggNOG-mapper (Huerta-Cepas et al., 2019) was used to annotate the Clusters of Orthologous Groups (COGs) classifications of accessory genes.

Serotyping and Multi-Locus Sequence Typing

In silico serotyping was performed using ECTyper[1] based on assembled genomes. MLST sequence type (MLST-ST) was obtained by scanning the sequences of seven house-keeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*) against PubMLST typing schemes using mlst.[2]

Virulence Factors and Plasmid Detection and Typing

The presence or absence of classic virulence genes (enterotoxins and CFs) and non-classic virulence genes were identified using ABRicate[3] and BLASTN by scanning against the ETEC virulence database[4] and a custom non-classic virulence genes database. We

[1] https://github.com/phac-nml/ecoli_serotyping
[2] https://github.com/tseemann/mlst
[3] https://github.com/tseemann/abricate
[4] https://github.com/avonm/ETEC_vir_db/

analyzed a total of 11 non-classic virulence genes, including *leoA* (accession number: AF170971), *tia* (U20318), *tibA* (AF109215), *etpA* (AY920525, position: 2718-8021), *cexE* (LR883053, position: 25424–25792), *clyA* (AY576657, position: 241–1152), *eatA* (AY163491), *yghJ* (DQ866820, position: 322–622), *irp2* (NC_003143, position: 2156049–2162156), *fyuA* (NC_003143.1, position: 2140840–2142861), and *astA* (AB042005). A gene was considered present if the overall coverage and identity were larger than 80%.

The ST and LT nucleotide sequences of Shenzhen isolates were extracted from assembled genomes using BLASTN and were aligned against the sequences of previously reported ST and LT variants[5] (Joffré et al., 2015, 2016; von Mentzer et al., 2021) using MAFFT v7.480 (Katoh and Standley, 2013). Multiple sequence alignments were used to construct the maximum likelihood phylogenetic trees, and the ST and LT variants of Shenzhen isolates were determined based on the phylogenetic trees (**Supplementary Figure 1**).

PlasmidFinder v2.1 (Carattoli et al., 2014) was used to detect and determine the incompatibility groups of plasmids based on assembled genomes with default settings.

Single-Nucleotide Polymorphism Calling and Phylogenetic Analyses

Core-genome (regions present in >99% strains) single-nucleotide polymorphisms (SNPs) were identified using Snippy v4.6.0 pipeline,[6] with *E. coli* K12 (accession number: NC_000913) as the reference genome. Briefly, for each isolate, clean sequencing reads were mapped to reference genome using BWA-mem (Li and Durbin, 2009), and SNPs were called using SAMtools (Li et al., 2009) and FreeBayes (Garrison and Marth, 2012). Repetitive regions of the reference genome were identified using Tandem Repeats Finder (TRF) (Benson, 1999) and by self-aligning using BLASTN as previously described (Yang et al., 2019). SNPs located in repetitive regions were removed prior to phylogenetic analysis. The maximum likelihood trees based on non-repetitive core-genome SNPs and alignments of ST and LT nucleotide sequences were constructed using IQ-TREE v2.0.3 (Minh et al., 2020) with auto-detected best-fitting substitution model, respectively.

Antimicrobial Resistance Mutation/Gene Detection and Phenotype Testing

ResFinder 4.1 (Bortolaia et al., 2020) was used to detect the AMR related mutations and genes based on clean sequencing reads with default settings. Based on the lineage classification and AMR mutation/gene distribution, we selected 52 representative isolates for AMR phenotype testing. The susceptibilities of the representative isolates were determined using disc diffusion assay against 17 antimicrobial agents, including NAL, AMP, TET, ampicillin/sulbactam (AMS), cefotaxime (CTX), trimethoprim/sulfamethoxazole (SXT), azithromycin (AZI), streptomycin (STR), ciprofloxacin (CIP),

[5] https://github.com/avonm/ETEC_toxin_variants_db
[6] https://github.com/tseemann/snippy

ceftazidime (CAZ), amikacin (AMK), chloramphenicol (CHL), colistin (CT), ceftazidime/avibactam (CZA), ertapenem (ETP), meropenem (MEM), and tigecycline (TGC). Breakpoints for sensitive, intermediate, and resistant were defined by the Clinical and Laboratory Standards Institute (CLSI) document M100-S26, and *E. coli* ATCC25922 was used as the control. ESBL-producing isolates were identified by assessing susceptibility phenotype to CTX, CAZ, and CZA.

RESULTS

Demographic and Clinical Symptom Characteristics

A total of 106 ETEC isolates from indigenous outpatients with diarrhea collected during routine foodborne disease surveillance between 2015 and 2020 were included in this study (**Supplementary Table 1**). These outpatients had a median age of 27 years (interquartile range: 21–36), of which only seven (7%) was children aged <5 years and 82 (77%) were adults aged >20 years. There were more males (51%, 54/106) than females (49%, 52/106), and the clinical symptoms included diarrhea (95%, 100/106), abdominal pain (29%, 31/106), vomiting (8%, 9/106), and fever (8%, 9/106).

Escherichia coli Phylogroup and Enterotoxigenic *Escherichia coli* Lineage

To investigate the *E. coli* phylogroup and ETEC lineage classification of Shenzhen isolates, we compared the 106 Shenzhen isolates to 71 representative isolates from the largest 50 *E. coli* lineages (Horesh et al., 2021) and 21 global ETEC lineages (von Mentzer et al., 2014), and constructed a maximum-likelihood (ML) tree of 177 isolates based on 267,822 SNPs in the core-genome (**Figure 1**). Shenzhen isolates showed a remarkably high diversity, which can be attributed to four *E. coli* phylogroups and 15 ETEC lineages (**Table 1** and **Figure 1**). The majority of Shenzhen isolates belonged to phylogroup A (71%, 75/106), and the remaining were assigned into phylogroup B1 (13%, 14/106), D (6%, 6/106), and E (10%, 11/106). Phylogroup A can be further divided into eight ETEC lineages, and lineage L11 (25%, 26/106), a novel lineage L2/4 (between L2 and L4, 21%, 22/106, **Figure 1**), and L4 (15%, 16/106) were major circulating lineages. Phylogroup B1 can be divided into five lineages (L5, L8, and L17–L19) each with few isolates ($n \leq 3$), and D and E each contained one lineage, a novel lineage L-N1 ($n = 6$) and L7 ($n = 11$), respectively (**Figure 1**).

Phylogroup A has always been dominant during the 6-year sampling period, while the fractions of other phylogroups were variable. At the end of sampling in 2020, only two phylogroups A and B1 persisted (**Figure 2A**). The fractions of different ETEC lineages were also variable. Lineage L11 was dominant before 2016; since 2017, multiple lineages coexisted and no obvious dominant lineage was identified (**Figure 2B**).

Serotype and Sequence Type

To compare with traditional subtyping results, we performed *in silico* serotyping and MLST-ST based on assembled genome sequences. There was a close association of ETEC lineage with serotype and MLST-ST, with isolates within a lineage belonging to 1–2 serotypes or STs (**Table 1**). The most prevalent serotype was O6:H16 (26%, 28/106), followed by O159:H34 (21%, 22/106) and O25:H16 (15%, 16/106). The most prevalent MLST-ST was ST48 (21%, 22/106), followed by ST218 (19%, 20/106) and ST1491 (15%, 16/106). The serotype and MLST-ST of three major lineages were O159:H34/O159:H43 and ST218/ST3153 for L11, O6:H16 and ST48 for L2/4, and O25:H16 and ST1491 for L4 (**Table 1**).

Virulence Factors

Heat-stable toxin (ST)-positive ETEC has been most prevalent (76%, 81/106) in Shenzhen throughout the sampling period (**Figure 2C**), followed by LT-positive ETEC (17%, 18/106). ST and LT-positive ETEC was only detected in seven isolates (7%) in three of the 6-year sampling period. Among ST-positive ETEC, STh-positive isolates (60%, 64/106) was more prevalent than STp-positive ones (16%, 17/106) throughout the sampling period (**Figure 2C**). More specifically, the most prevalent ST variant was STa3/4 (57%, 60/106), followed by STa5 (15%, 16/106) and STa7 (8%, 9/106); the most prevalent LT variant was LT17 (15%, 16/106). Two novel variants of STp and LT, designated as STa8 and LT29, were identified in one isolate, respectively (**Table 1**, **Supplementary Table 1**, and **Supplementary Figure 1**).

One to three known CFs were identified in 68 (64%) Shenzhen isolates. The most prevalent CF was CS21 (48%, 51/106, with or without other CFs), followed by CS6 (34%, 36/106) throughout the sampling period (**Figure 2D**). Except CS21 and CS6, the fraction of other CFs was low and they were only detected in 3 years. The most prevalent combination was CS6 + CS21 (18%, 19/106).
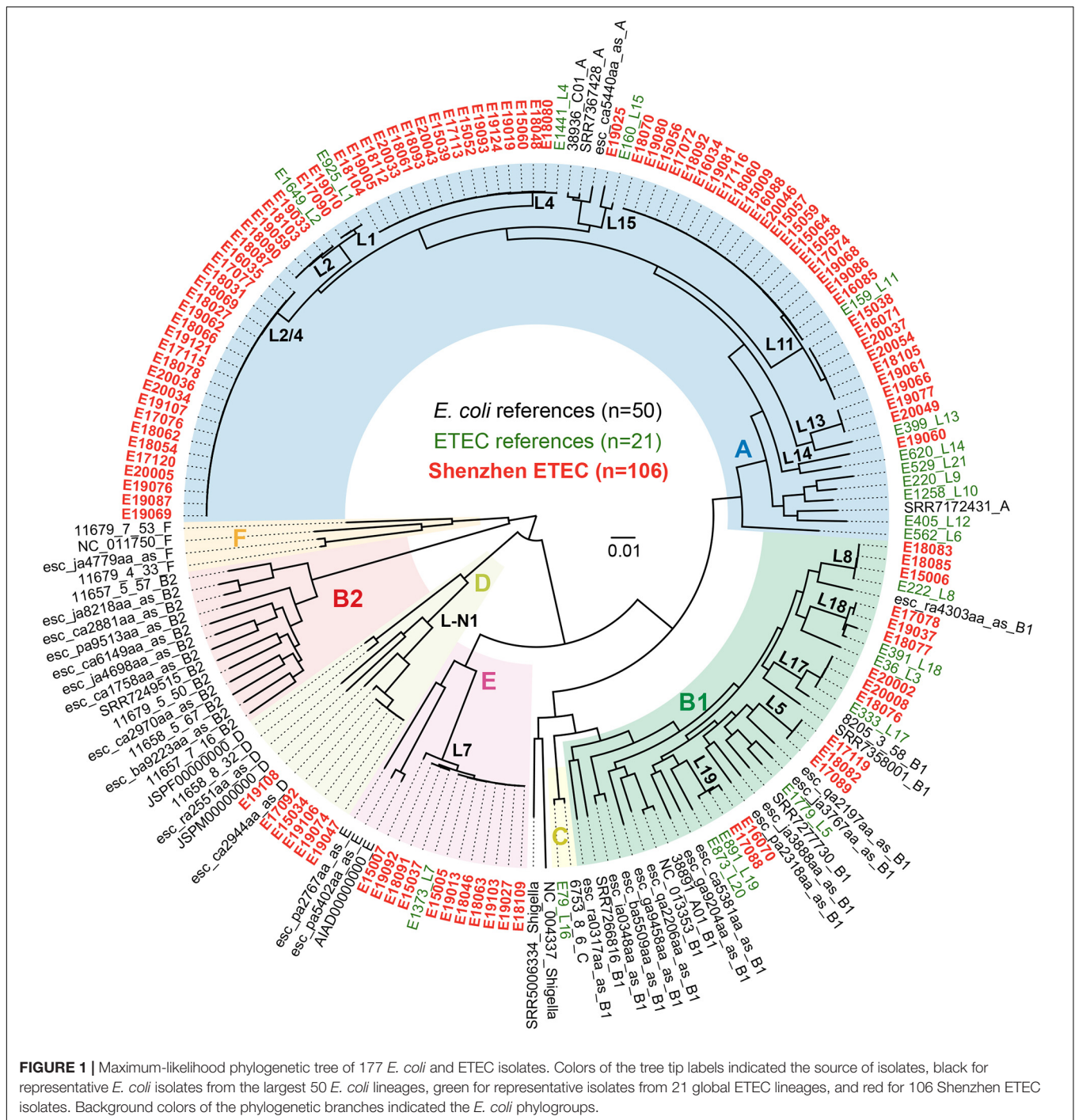
In addition to enterotoxins and CFs, 10 of 11 detected non-classic virulence genes were identified in Shenzhen isolates with different frequencies (2–100%, **Table 1** and **Supplementary Table 1**). The hemolysin gene *clyA* (100%, 106/106) and mucinase gene *yghJ* (97%, 103/106) were most prevalent in Shenzhen isolates, followed by enteroaggregative heat-stable toxin gene *astA* (75%, 79/106), adhesion gene *etpA* (63%, 67/106), and another mucinase gene *eatA* (30%, 32/106). The other five non-classic virulence genes were detected with low frequencies (2–8%). Co-presence of classic and non-classic virulence genes was identified. The *etpA* (94%, 60/64) and *astA* (91%, 58/64) genes were mostly detected along with STh, and *eatA* (89%, 16/18) gene was mostly detected along with LT.

The distribution of classic and non-classic virulence genes in different lineages was shown in **Table 1** and **Supplementary Table 1**. Twelve (80%) of 15 lineages were ST-positive (53%, 8/15) or contained ST-positive isolates (27%, 4/15), and STh-positive lineages (40%, 6/15) or lineages containing STh-positive isolates (20%, 3/15) were more prevalent than STp-positive lineages (13%, 2/15) or lineages containing STp-positive isolates (13%, 2/15). CFs were detected in 11 lineages (73%), and the most prevalent CFs, CS21, and CS6 were detected in seven (47%) and

TABLE 1 | Classification, enterotoxin, and colonization factor distribution of 106 ETEC isolates in Shenzhen, 2015-2020.

E. coli phylogroup	ETEC lineage	No. of isolates	Serotype[1]	Sequence type (ST)[1]	Enterotoxin variant	Colonization factor	Non-classic virulence genes	Plasmid incompatibility groups
A (n = 75)	L1	2	O6:H16	ST6955	STh (STa3/4) + LT (LT20)	CS1 + CS3 + CS21	*tia*, *etpA*, *cexE*, *clyA*, *eatA*, *yghJ*, *astA*	FII, FIB, I1 (n = 1); FII, FIB, I1, B/O/K/Z (n = 1)
	L2	4	O6:H16	ST4	STh (STa3/4) + LT (LT20)	CS2 + CS3 + CS21	*etpA*, *clyA*, *eatA*, *yghJ*, *astA*	FII, FIB (n = 2); FII, FIB, I2 (n = 3)
	L2/4	22	O6:H16	ST48	STh (STa3/4, n = 21); STp (STa1, n = 1)	CS21 (n = 21)	*etpA*, *clyA*, *yghJ*, *astA* (n = 21); *etpA*, *clyA*, *eatA*, *yghJ*, *astA* (n = 1)	FII, FIB, I1 (n = 18); FII, FIB (n = 1); FII, FIB, I1, B/O/K/Z (n = 1); FII, I1, N (n = 1)
	L4	16	O25:H16	ST1491	LT (LT17)	CS6 + CS21	*clyA*, *eatA*, *yghJ*	FII, FIB (n = 4); FII, FIB, B/O/K/Z (n = 6); FII, FIB, B/O/K/Z, Y (n = 1); FII, FIB, I1 (n = 1)
	L11	26	O159:H34 (n = 20); O159:H43 (n = 6)	ST218 (n = 20), ST3153 (n = 6)	STh (STa3/4)		*etpA*, *clyA*, *yghJ*, *astA*	FII (n = 13); FII, B/O/K/Z (n = 7); FII, I1 (n = 6)
	L13	3	O-:H10	ST226	STh (STa3/4)		*etpA*, *clyA*, *yghJ*, *astA*	FII, I1
	L14	1	O25:H42	ST1201	STh (STa2)	CS4 + CS6	*clyA*, *eatA*, *yghJ*, *astA*	FII, FIB, B/O/K/Z
	L15	1	O107/117:H27	ST10	STh (STa7)	CS21	*etpA*, *clyA*, *eatA*, *yghJ*	FII, FIB
B1 (n = 14)	L5	3	O159:H20	ST2040	STh (STa3/4)		*etpA*, *clyA*, *yghJ*, *astA*	FII, B/O/K/Z
	L8	3	O148:H28	ST94	STh (STa7)	CS6 + CS21	*clyA*, *astA*	FII, FIB
	L17	3	O27:H7	ST316	STp (STa5)	CS6	*clyA*, *yghJ*	FII, FIB
	L18	3	O159:H34 (n = 2); O23:H16 (n = 1)	ST1490 (n = 2), ST- (n = 1)	STp (n = 2, STa5), LT (n = 1, LT29)	CS6 + CS23 (n = 2)	*clyA*, *eatA*, *yghJ*, *astA*	FII, FIB
	L19	2	O29:H12 (n = 1); O153:H12 (n = 1)	ST155	STh (n = 1, STa3/4), LT (n = 1, LT15)		*etpA*, *clyA*, *yghJ*, *astA* (n = 1); *clyA*, *yghJ* (n = 1)	FII (n = 1); FII, FIB (n = 1)
D (n = 6)	L-N1	6	O15:H18 (n = 5); O-/H18 (n = 1)	ST69 (n = 5), ST1380 (n = 1)	STh (n = 5, STa7), STp + LT (STa8 + LT4, n = 1)	CS21 (n = 4)	*etpA*, *clyA*, *eatA*, *yghJ*, *irp2*, *fyuA* (n = 5); *clyA*, *yghJ*, *astA* (n = 1)	FII, FIB (n = 4); FII (n = 1); FII, FIB, B/O/K/Z, FIA (n = 1)
E (n = 11)	L7	11	O169:H41 (n = 10); O167:H41 (n = 1)	ST182 (n = 10), ST3930 (n = 1)	STp (STa5)	CS6	*leoA*, *tia*, *clyA*, *yghJ*, *astA* (n = 7); *clyA*, *yghJ*, *astA* (n = 4)	FII, FIB

[1]Untypeable serotype O and MLST sequence type was denoted by O- and ST-, respectively.

FIGURE 1 | Maximum-likelihood phylogenetic tree of 177 *E. coli* and ETEC isolates. Colors of the tree tip labels indicated the source of isolates, black for representative *E. coli* isolates from the largest 50 *E. coli* lineages, green for representative isolates from 21 global ETEC lineages, and red for 106 Shenzhen ETEC isolates. Background colors of the phylogenetic branches indicated the *E. coli* phylogroups.
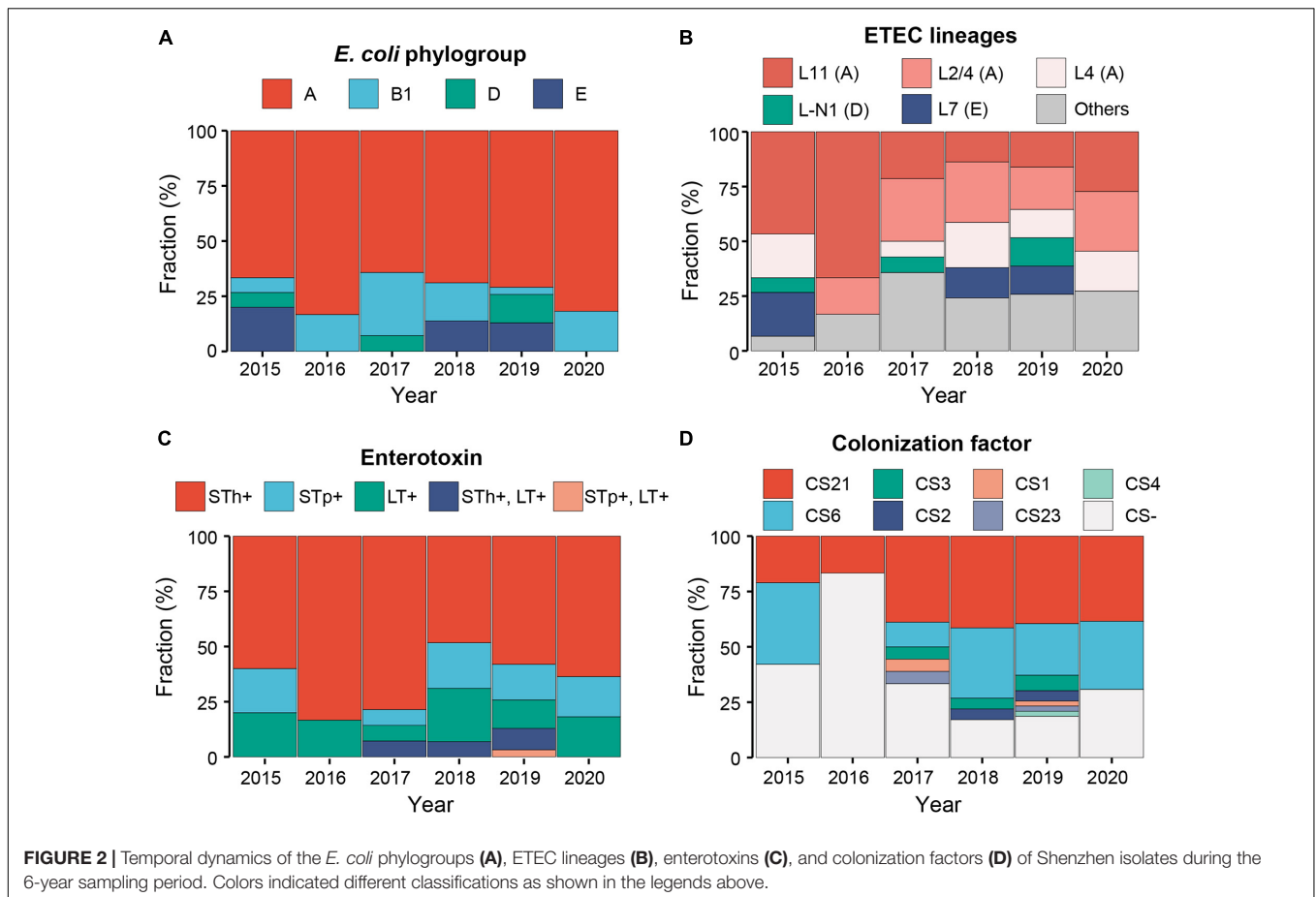
six (40%) lineages, respectively. Except two prevalent non-classic virulence genes *clyA* and *yghJ*, *astA*, *etpA*, and *eatA* genes were detected in 12 (80%), 9 (60%), and 8 (53%) lineages, respectively.

In addition to the association between lineage and serotype and MLST-ST, the association of lineage with enterotoxin and CF was also identified. Most isolates within a lineage shared identical virulence gene (classic and non-classic) profiles, except for five lineages (L2/4, L7, L18, L19, and L-N1) in which two or three types of virulence gene combinations were detected. In

contrast, the virulence gene profiles of different lineages were largely distinct (**Table 1** and **Supplementary Table 1**).

Plasmid Incompatibility Groups

In silico plasmid typing analysis identified replicons belonging to eight incompatibility (Inc) groups in all the 106 Shenzhen isolates, with one to four Inc groups present in each isolate (**Table 1** and **Supplementary Table 1**). The replicon IncFII was most prevalent and can be detected in all Shenzhen isolates,

FIGURE 2 | Temporal dynamics of the *E. coli* phylogroups **(A)**, ETEC lineages **(B)**, enterotoxins **(C)**, and colonization factors **(D)** of Shenzhen isolates during the 6-year sampling period. Colors indicated different classifications as shown in the legends above.

followed by IncFIB (67%, 71/106), IncI1 (31%, 33/106), and IncB/O/K/Z (20%, 21/106); the other replicons (IncFIA, IncI2, LncN, and LncY) were only detected in one isolate, respectively. There was an association between the replication IncFIB and virulence gene CS21, and 72% (51/71) IncFII-positive isolates also encode CS21. Only one type of replicon profile was identified in eight lineages, and the other seven lineages each contained two to four types of replicon profiles (**Table 1** and **Supplementary Table 1**).

Pan-Genome Analysis

We performed pan-genome analysis of 127 ETEC genomes of Shenzhen isolates ($n = 106$) and globally representative isolates ($n = 21$) to identify genes unique to or significantly associated with Shenzhen isolates. A total of 11,813 pan-genes were identified in these 127 genomes, including 3,307 core-genes (present in $\geq$99% isolates) and 8,506 accessory genes. There were no genes unique to Shenzhen or non-Shenzhen isolates; however, we identified significant differences ($p < 0.01$, Fisher's exact tests) in the frequencies of 589 accessory genes between Shenzhen and non-Shenzhen isolates (**Supplementary Table 2**). There were 178 accessory genes significantly enriched in Shenzhen isolates, of which two were known virulence genes STh and *etpA*, and the other genes encoded hypothetical proteins with unknown functions (46%) or were associated

with replication and repair (COG category L; 37%) and cell wall/membrane/envelop biogenesis (COG category M; 19%). For example, STh was present in 66% Shenzhen isolates and 29% non-Shenzhen isolates, respectively; the gene *group_4545* encoding a hypothetical protein was present in 91% Shenzhen isolates but in only 19% non-Shenzhen isolates. In addition, there were 411 accessory genes significantly enriched in non-Shenzhen isolates including the virulence gene LT (**Supplementary Table 2**).

We further attempted to identify genes unique to two novel lineages L2/4 and L-N1. A total of 25 L2/4 lineage unique genes were identified (**Supplementary Table 3**), of which 76% encoded hypothetical proteins and 20% were associated with replication and repair (COG category L). There were 105 genes unique to L-N1 lineage (**Supplementary Table 3**), of which 72% encoded hypothetical proteins, 10% were associated with cell motility (COG category N) and intracellular trafficking and secretion (COG category U), and 5% were associated with transcription (COG category K). None of these L2/4 and L-N1 lineage unique genes were known ETEC virulence genes.

Antimicrobial Susceptibility

To characterize the AMR profiles of Shenzhen isolates, we firstly scanned the genomes to identify AMR related mutations and genes. A total of 21 AMR mutations/genes were detected (**Figure 3**), which were involved in resistance to

five classes of antimicrobial agents, including fluoroquinolone (two mutations and two genes), beta-lactam (seven genes), tetracycline (two genes), macrolide (one gene), and folate pathway antagonist (seven genes).

Fluoroquinolone resistance mutations/genes were most prevalent, which were detected in 99 (93%) isolates. For example, DNA gyrase gene *gyrA* p.S83L mutation and quinolone resistance gene *qnrS1* were detected in 63 (59%) and 37 (35%) isolates, respectively. Beta-lactam resistance genes ranked second and were detected in 63 (59%) isolates, of which bla_{TEM-1B} and $bla_{CTX-M-15}$ were more prevalent and were detected in 40 (38%) and 18 (17%) isolates, respectively. Tetracycline resistance genes were detected in 34 (32%) isolates, of which *tet(A)* and *tet(B)* were detected in 24 (23%) and 11 (10%) isolates, respectively. A macrolide resistance gene *mph(A)* was detected in 29 (27%) isolates. Notably, this gene has been reported to mediate azithromycin resistance, which is usually less frequent in ETEC (Xiang et al., 2020). Folate pathway antagonist resistance genes, including *sul1, sul2, dfrA1, dfrA8, dfrA14, dfrA15,* and *dfrA17* were detected in 22 (21%) isolates, and each gene appeared in 3–13 isolates.

Unlike the close association of lineage with serotype, MLST-ST and virulence factors, there was no obvious association between lineages and most AMR mutations/genes, with multiple AMR mutations/genes combinations being identified in isolates of a lineage (**Figure 3**). Notably, most isolates of two major lineages L11 and L2/4 had *gyrA* p.S83L mutations, and all the L2/4 lineage isolates also carried bla_{TEM-1B} and *mph(A)* genes.

We further selected 52 isolates representing all the lineages and AMR mutations/genes combinations for AMR phenotype testing against 17 antimicrobial agents (**Table 2** and **Figure 3**). AMR mutations/genes distributions were generally concordant with phenotype testing results. The resistance to a quinolone agent, NAL was highest (71%, 37/52) in Shenzhen isolates, and 78% (29/37) of them carried *gyrA* (p.S83L or p.S83A) mutations. However, only two (4%) isolates were resistant to another quinolone agent ciprofloxacin. The resistance to beta-lactam agents, including AMP (69%, 36/52), AMS (46%, 24/52), and CTX (31%, 16/52) was also prevalent. Sixteen (31%) representative isolates were identified as ESBL-producing, of which 81% (13/16) encode at least one type of β-lactamase associated gene, with $bla_{CTX-M-15}$ (50%, 8/16) and $bla_{CTX-M-14}$ (25%, 4/16) being most common. In addition, the resistance to other three classes of antimicrobial agents were detected, including AZI (38%, 20/52, macrolide class), TET (37%, 19/52, tetracycline class), and SXT (29%, 15/52, folate pathway antagonist class). There was a close association between azithromycin resistance and the presence of *mph(A)* gene, with 18 of 20 azithromycin-resistant isolates encoding *mph(A)*. No resistance to seven antimicrobial agents was detected, including AMK, CHL, CT, CZA, ETP, MEM, and TGC. All (12/12) the tested representative L2/4 lineage isolates were resistant to at least three classes of agents, i.e., multidrug-resistant (MDR); 16 (40%) of 40 representative isolates of non-L2/4 lineages were MDR (**Table 2** and **Figure 3**).

DISCUSSION

Despite the high prevalence in children and travelers, ETEC also lead to substantial infections in adults in endemic areas (Lamberti et al., 2014). Moreover, it is intriguing that most ETEC infections in Shenzhen (Li et al., 2017), as well as several other developed cities and province in China (Chen Y. et al., 2014; Qu et al., 2014; Pan et al., 2015), such as Beijing and Shanghai, were from indigenous adults. However, there is currently a lack of understanding on the epidemiology of ETEC infections in indigenous adults. Compared to traditional molecular subtyping methods, WGS not only provides the ultimate resolution on reconstruction the relationships between isolates, but also can be used for secondary in-depth virulence factors and AMR genes analysis, which has been one of the most powerful methods to characterize the epidemiology of pathogens (Didelot et al., 2012; Allard et al., 2016; Ronholm et al., 2016). We reconstructed the WGS-based population structure of Shenzhen isolates in the context of global ETEC and *E. coli* lineages, and linked the genomic diversity to traditional subtyping results based on serotyping and MLST-ST. In addition, we characterized the virulence factors, AMR mutation/gene and phenotype profiles of Shenzhen isolates. To our knowledge, this is the first study on the genomic epidemiology and AMR profile of ETEC infections in indigenous adults.

Shenzhen ETEC isolates showed a remarkable diversity, which can be attributed to four *E. coli* phylogroups, with the majority falling into phylogroup A (71%). The phylogroup distribution is different from that of global (von Mentzer et al., 2014) and Bangladeshi (Sahl et al., 2017) isolates [high fraction (>40%) of B1 isolates], and is similar to that of Chilean isolates (phylogroup A: 81%) (Rasko et al., 2019). More specifically, three major lineages, L11 (25%), L2/4 (21%), and L4 (15%), persisted in Shenzhen throughout the 6-year sampling period, together accounting for 60% cases. L11 can be detected in children aged <5 years and adult travelers from multiple countries, but at a low fraction (4%) in the global dataset (von Mentzer et al., 2014). If taken serotype and MLST-ST into consideration, L11 (O159:H34, ST218) isolates was only identified in local diarrheal cases of Korea (Chung et al., 2019) and China (Chen Y. et al., 2014; Pan et al., 2015; Li et al., 2017), and the virulence factor profiles (STh+, without known CS) of isolates from two countries were identical, indicating the possibility of recent transmission. L2/4 was a previously undefined lineage, which located between L2 and L4 in the phylogenic tree. However, it had a different virulence factor profile (STh: 21/22, STp: 1/22, CS21: 22/22) from L2 (global dataset: STh + LT, CS2 + CS3 ± CS21) or L4 (global dataset: STh/LT, CS6 ± CS21/CS6 + CS8/CS21). The serotype and MLST-ST of L2/4 was O6:H16 and ST48, respectively, and ETEC isolates with this subtyping combination were previously identified from pigs in Denmark (García et al., 2020) and from human cases in China (Chen Y. et al., 2014; Li et al., 2017). However, no CF was reported in Danish isolates whereas all Shenzhen isolates carried CS21. L4 lineage was also detected at low fraction (6%) in the global dataset (von Mentzer et al., 2014), and the subtyping combination of Shenzhen isolates, i.e., L4 (O25:H16, ST1491), was only identified in isolates from few travelers returning to the
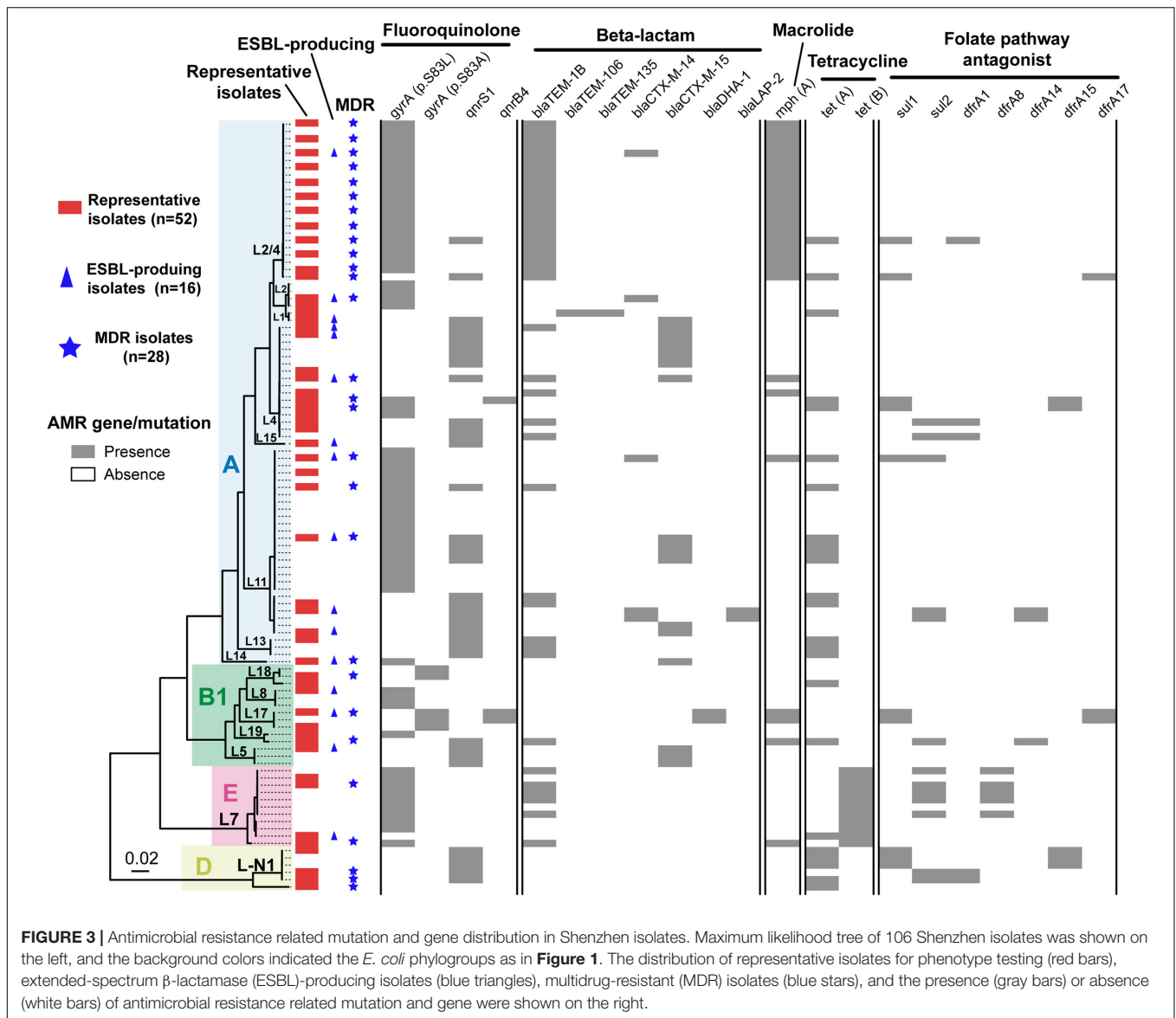
FIGURE 3 | Antimicrobial resistance related mutation and gene distribution in Shenzhen isolates. Maximum likelihood tree of 106 Shenzhen isolates was shown on the left, and the background colors indicated the *E. coli* phylogroups as in **Figure 1**. The distribution of representative isolates for phenotype testing (red bars), extended-spectrum β-lactamase (ESBL)-producing isolates (blue triangles), multidrug-resistant (MDR) isolates (blue stars), and the presence (gray bars) or absence (white bars) of antimicrobial resistance related mutation and gene were shown on the right.

UK (Boxall et al., 2020) and from local diarrheal cases of China (Chen Y. et al., 2014; Pan et al., 2015; Li et al., 2017). Taken together, these results revealed that the major circulating ETEC lineages in Shenzhen, as well as in several other developed regions of China were distinctive, which are not the commonly detected global or endemic lineages.

In addition to subtyping, the virulence factor profile of Shenzhen and other Chinese ETEC isolates was also distinctive. ST-positive ETEC has been dominant (76%) in Shenzhen throughout the sampling period and 80% circulating lineages were ST-positive (53%) or contained ST-positive isolates (27%), whereas the fractions of LT-positive and ST + LT-positive isolates and lineages were lower. CS21 was the most prevalent CF (48%) in Shenzhen isolates, followed by CS6 (34%). Similar virulence factor profile was also observed in ETEC isolates from diarrheal cases in Shanghai, China (ST: 74%, CS21: 63%,

CS6: 41%) (Pan et al., 2015). In contrast, a systematic review showed that globally, ST-positive, LT-positive, and ST + LT-positive ETEC accounted for 50, 25, and 27% non-travel human infections, respectively, and the most prevalent CFs were CFA/I, followed by CS21 and CS6 (Isidean et al., 2011). There was a considerable variability of virulence factor profiles across regions and populations. High ST-prevalence was only observed among travelers in East Asia/Pacific (75%) and non-travelers in Europe/Central Asia (77%), however, the most prevalent CF in East Asia/Pacific was CFA/I and CS21 was rarely detected (CF-prevalence in Europe/Central Asia is unavailable) (Isidean et al., 2011). Interestingly, CS21 was most prevalent CFs in ETEC-endemic regions including Latin America/Caribbean and the Middle East/North Africa, and among global travelers with a frequency of ∼22% (Isidean et al., 2011). Notably, the most commonly detected CF globally, CFA/I, was not identified in

TABLE 2 | Antimicrobial resistance of 52 representative Shenzhen ETEC isolates.

Antimicrobial agent[1]	Antimicrobial resistance rate (%)						
	L11 (*n* = 7)	L2/4 (*n* = 12)	L4 (*n* = 10)	L7 (*n* = 4)	L-N1 (*n* = 4)	Other[2] (*n* = 15)	Total (*n* = 52)
NAL	71	92	60	75	50	67	71
AMP	86	100	40	75	25	67	69
AMS	29	100	20	25	0	47	46
AZI	14	100	20	25	0	27	38
TET	43	50	20	100	100	27	37
CTX	57	25	30	25	0	47	31
SXT	14	50	30	25	75	33	29
STR	43	25	0	25	50	20	19
CIP	0	0	10	0	0	7	4
CAZ	0	0	0	0	0	7	2
ESBL	57	8	30	25	0	47	31
MDR	43	100	30	50	75	33	54

[1]Only 10 of 17 antimicrobial agents with at least one resistant isolate were shown. [2]Lineages with <3 isolates together were defined as others. NAL, nalidixic acid; AMP, ampicillin; TET, tetracycline; AMS, ampicillin/sulbactam; CTX, cefotaxime; SXT, trimethoprim/sulfamethoxazole; AZI, azithromycin; STR, streptomycin; CIP, ciprofloxacin; CAZ, ceftazidime. ESBL, extended-spectrum β-lactamase; MDR, multidrug-resistant, defined as resistant to ≥3 classes of antimicrobial agents.

Shenzhen isolates. Moreover, a recent Global Enteric Multicenter Study (GEMS) report (Vidal et al., 2019) showed that the prevalence of ST, LT, and ST + LT among ETEC isolates from children aged <5 years with moderate-to-severe diarrhea in Africa and Asia were 36, 32, and 32%, respectively; the prevalence from the matched controls were 21, 47, and 33%, respectively. The most commonly detected CFs were CFA/IV (CS6 alone or with CS4 or CS5), CS5 + CS6, and CFA/I. Among all the sampling sites of this study, ST-positive ETEC isolates were not dominant (<50%), and CS21 was rarely detected (<2%). More specifically, the most prevalent ST variants in Shenzhen isolates were STa3/4 and STa5, which were consistent with that in global isolates (Joffré et al., 2016). However, the most prevalent LT variant in Shenzhen isolates was LT17. By contrast, the most prevalent LT variant in global isolates was LT1 (41%), while LT17 was rarely detected (2.1%) (Joffré et al., 2015).

The unique ETEC epidemiology in Shenzhen and China, i.e., most infections were from indigenous adults, might be related to several reasons. First, the pathogenicity of China ETEC isolates may be higher in adults than in children, given the distinctive lineage and virulence factor profiles. This hypothesis was supported by the observation in Guatemala (Torres et al., 2015), where ST-positive ETEC infections were significantly more prevalent in adult travelers compared to indigenous children, suggesting higher pathogenicity of ST-positive ETEC in adults, whereas most Chinese ETEC isolates were ST-positive. Moreover, a recent study showed that a ST enterotoxin variant STa5 was associated with ETEC infections in adults, suggesting that specific type of ETEC may have higher pathogenicity in adults (Joffré et al., 2016). However, the prevalence of STa5 variant in Shenzhen isolates was only 15%, indicating the existence of other known/unknown virulence genes associated with ETEC infections in adults. We identified multiple accessory genes significantly enriched in Shenzhen isolates by pan-genome analysis including the known virulence genes STh and *etpA*, which provides candidate targets for further

studies. Second, different characterization between children and adult populations, such as eating habits and immunity, may also be associated with the unique ETEC epidemiology. For example, a study showed that dietary calcium can improve human resistance to ETEC infection (Bovee-Oudenhoven et al., 2003), while children usually tend to consume more calcium than the adult population. Besides, it has been reported that breastfeeding may protect infants against severe ETEC infection (Clemens et al., 1997). Furthermore, eating out is a major risk factor of foodborne disease (Liu et al., 2018), and the frequency in the adult population is usually higher than that in children. In addition, the immunity provided by previous ETEC infection or vaccine may decrease over age, leading to more ETEC infections in adults.

Characterizing the antibiotic susceptibility profile of ETEC would be helpful to guide the clinical treatment. In this study, we investigated the AMR mutation/genes and phenotype of Shenzhen ETEC isolates by integrating WGS-based analysis and antimicrobial susceptibility testing. AMR mutation/gene profile was generally concordant with the phenotype testing results of 52 representative isolates, which revealed high resistance rate to NAL (71%), AMP (69%), and AMS (46%). In recent years, high resistance of ETEC to these commonly used agents was also reported in multiple other countries including Peru (Rivera et al., 2010), Bangladesh (Begum et al., 2016), South Korea (Oh et al., 2014), and in other cities of China (Chen Y. et al., 2014; Pan et al., 2015). Due to the increased AMR, new antimicrobial agents such as azithromycin have been used as the first-line agent for ETEC infection treatment. Azithromycin is a broad-spectrum macrolide antimicrobial agent against several bacterial species, and is very effective for *Enterobacteriaceae* infection treatment (Gomes et al., 2019). However, azithromycin-resistant ETEC isolates from diarrheal patients have recently been reported in several countries at a moderate frequency (10–30%) (Begum et al., 2016; Guiral et al., 2019), and in Shanghai, China at a very high frequency (87%) (Xiang et al., 2020). Macrolides inactivation mediated by macrolide-resistant phosphotransferase *mph(A)*

gene was the most common mechanism for the azithromycin resistance (Gomes et al., 2019; Xiang et al., 2020). We found that 38% representative Shenzhen ETEC isolates were resistant to azithromycin, and most of these isolates carried *mph(A)* gene. Moreover, we showed that the prevalence of ESBL-producing ETEC isolates from outpatients was 31% in Shenzhen, China, which is similar to that of diarrheal patients in Nepal post-2013 (30%) (Margulieux et al., 2018) and travel cases to Southeast Asia/India (43%) (Guiral et al., 2019). In addition, nearly all the isolates of a major circulating lineage L2/4 carried *gyrA* p.S83L mutation, bla_{TEM-1B} and *mph(A)* genes, and AMR phenotype testing showed that all the representative L2/4 isolates were resistant to AMP, AMS, and AZI and most (11/12) of them were resistant to NAL, suggesting that L2/4 was a MDR lineage. The identification of azithromycin-resistant, ESBL-producing ETEC and a major MDR lineage L2/4 in Shenzhen highlighted the importance of ongoing AMR surveillance.

In summary, during routine foodborne disease surveillance, we found that the epidemiology of ETEC infections in Shenzhen, China is distinctive, with most infections occurring in indigenous adults. By integrating WGS and antimicrobial susceptibility testing, we characterized the temporal dynamics of population structure and virulence factors, and the AMR mutation/gene and phenotype profile of Shenzhen ETEC isolates in 6 years. Shenzhen ETEC isolates showed a remarkable high diversity, which belonged to four *E. coli* phylogroups and 15 ETEC lineages, and the major virulence factors were enterotoxin ST and CF CS21 and CS6. Intriguingly, the major circulating lineages in Shenzhen and their virulence factor profiles were distinctive, which are different from the commonly detected global or endemic ETEC lineages. Furthermore, we showed that AMR mutation/gene profiles of genomes were concordant with the phenotype testing results, and revealed that Shenzhen isolates had high resistance rates to several commonly used antimicrobial agents and identified a MDR lineage. To our knowledge, our study provides novel insight into the genomic epidemiology and antimicrobial susceptibility profile of ETEC infections in indigenous adults for the first time, which will not only enhance our comprehensive understanding on ETEC epidemiology, but also have implications for the development of vaccine and future surveillance and prevention of ETEC infections.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

CY, YC, CW, and QH conceived and designed the study. CY, YL, LZ, MJ, LX, ML, YS, LW, YJ, SW, RC, and XS performed the experiments and data analysis. CY and LZ wrote the original draft. CY, XZ, YL, MJ, YC, CW, and QH reviewed and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.732068/full#supplementary-material

REFERENCES

Abraham, S., Trott, D. J., Jordan, D., Gordon, D. M., Groves, M. D., Fairbrother, J. M., et al. (2014). Phylogenetic and molecular insights into the evolution of multidrug-resistant porcine enterotoxigenic *Escherichia coli* in Australia. *Int. J. Antimicrob. Agents* 44, 105–111. doi: 10.1016/j.ijantimicag.2014.04.011

Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., et al. (2016). Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J. Clin. Microbiol.* 54, 1975–1983. doi: 10.1128/JCM.00081-16

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Begum, Y. A., Talukder, K. A., Azmi, I. J., Shahnaij, M., Sheikh, A., Sharmin, S., et al. (2016). Resistance pattern and molecular characterization of enterotoxigenic *Escherichia coli* (ETEC) strains isolated in Bangladesh. *PLoS One* 11:e0157415. doi: 10.1371/journal.pone.0157415

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* 75, 3491–3500. doi: 10.1093/jac/dkaa345

Bovee-Oudenhoven, I. M., Lettink-Wissink, M. L., Van Doesburg, W., Witteman, B. J., and Van Der Meer, R. (2003). Diarrhea caused by enterotoxigenic *Escherichia coli* infection of humans is inhibited by dietary calcium. *Gastroenterology* 125, 469–476. doi: 10.1016/S0016-5085(03)00884-9

Boxall, M. D., Day, M. R., Greig, D. R., and Jenkins, C. (2020). Antimicrobial resistance profiles of diarrhoeagenic *Escherichia coli* isolated from travellers returning to the UK, 2015-2017. *J. Med. Microbiol.* 69, 932–943. doi: 10.1099/JMM.0.001214

Carattoli, A., Zankari, E., Garciá-Fernández, A., Larsen, M. V., Lund, O., Villa, L., et al. (2014). In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Chen, Q., Shi, X., Li, Y., Jiang, Y., Lin, Y., Qiu, Y., et al. (2014). Rapid genetic typing of diarrheagenic *Escherichia coli* using a two-tube modified molecular beacon based multiplex real-time PCR assay and its clinical application. *Ann. Clin. Microbiol. Antimicrob.* 13, 1–9.

Chen, Y., Chen, X., Zheng, S., Yu, F., Kong, H., Yang, Q., et al. (2014). Serotypes, genotypes and antimicrobial resistance patterns of human diarrhoeagenic *Escherichia coli* isolates circulating in southeastern China. *Clin. Microbiol. Infect.* 20, 52–58. doi: 10.1111/1469-0691.12188

Chung, S. Y., Kwon, T., Bak, Y. S., Park, J. J., Kim, C. H., Cho, S. H., et al. (2019). Comparative genomic analysis of enterotoxigenic *Escherichia coli* O159 strains isolated from diarrheal patients in Korea. *Gut Pathog.* 11, 1–10. doi: 10.1186/s13099-019-0289-6

Clemens, J. D., Rao, M. R., Chakraborty, J., Yunus, M., Ali, M., Kay, B., et al. (1997). Breastfeeding and the risk of life-threatening enterotoxigenic*Escherichia coli* diarrhea in Bangladeshi infants and children. *Pediatrics* 100:e2.

Del Canto, F., Valenzuela, P., Cantero, L., Bronstein, J., Blanco, J. E., Blanco, J., et al. (2011). Distribution of classical and nonclassical virulence genes in enterotoxigenic *Escherichia coli* isolates from Chilean children and tRNA gene screening for putative insertion sites for genomic islands. *J. Clin. Microbiol.* 49, 3198–3203. doi: 10.1128/JCM.02473-10

Denamur, E., Clermont, O., Bonacorsi, S., and Gordon, D. (2021). The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 19, 37–54. doi: 10.1038/s41579-020-0416-x

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13, 601–612. doi: 10.1038/nrg3226

Fleckenstein, J. M., and Kuhlmann, F. M. (2019). Enterotoxigenic *Escherichia coli* Infections. *Curr. Infect. Dis. Rep.* 21:9. doi: 10.1007/s11908-019-0665-x

Gaastra, W., and Svennerholm, A.-M. (1996). Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol.* 4, 444–452.

García, V., Gambino, M., Pedersen, K., Haugegaard, S., Olsen, J. E., and Herrero-Fresno, A. (2020). F4- and F18-Positive Enterotoxigenic *Escherichia coli* Isolates from Diarrhea of Postweaning Pigs: genomic Characterization. *Appl. Environ. Microbiol.* 86, 1–18. doi: 10.1128/AEM.01913-20

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* [Preprint]. arXiv:1207.3907.

Gomes, C., Ruiz-Roldán, L., Mateu, J., Ochoa, T. J., and Ruiz, J. (2019). Azithromycin resistance levels and mechanisms in *Escherichia coli*. *Sci. Rep.* 9:6089. doi: 10.1038/s41598-019-42423-3

Gonzales, L., Sanchez, S., Zambrana, S., Iñiguez, V., Wiklund, G., Svennerholm, A. M., et al. (2013). Molecular characterization of enterotoxigenic *escherichia coli* isolates recovered from children with diarrhea during a 4-year period (2007 to 2010) in bolivia. *J. Clin. Microbiol.* 51, 1219–1225. doi: 10.1128/JCM.02971-12

Guiral, E., Quiles, M. G., Muñoz, L., Moreno-Morales, J., Alejo-Cancho, I., Salvador, P., et al. (2019). Emergence of Resistance to Quinolones and -Lactam Antibiotics in Enteroaggregative and Enterotoxigenic *Escherichia coli* Causing Traveler's Diarrhea. *Antimicrob. Agents Chemother.* 63, 1–9. doi: 10.1128/AAC.01745-18

Horesh, G., Blackwell, G. A., Tonkin-Hill, G., Corander, J., Heinz, E., and Thomson, N. R. (2021).). A comprehensive and high-quality collection of *Escherichia coli* genomes and their genes. *Microb. Genomics* 7:000499. doi: 10.1099/mgen.0.000499

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Isidean, S. D., Riddle, M. S., Savarino, S. J., and Porter, C. K. (2011). A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine* 29, 6167–6178. doi: 10.1016/j.vaccine.2011.06.084

Joffré, E., von Mentzer, A., Abd El Ghany, M., Oezguen, N., Savidge, T., Dougan, G., et al. (2015). Allele variants of enterotoxigenic *Escherichia coli* heat-labile toxin are globally transmitted and associated with colonization factors. *J. Bacteriol.* 197, 392–403. doi: 10.1128/JB.02050-14

Joffré, E., von Mentzer, A., Svennerholm, A. M., and Sjöling, Å (2016). Identification of new heat-stable (STa) enterotoxin allele variants produced by human enterotoxigenic *Escherichia coli* (ETEC). *Int. J. Med. Microbiol.* 306, 586–594. doi: 10.1016/j.ijmm.2016.05.016

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

Khalil, I. A., Troeger, C., Blacker, B. F., Rao, P. C., Brown, A., Atherly, D. E., et al. (2018). Morbidity and mortality due to *shigella* and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* 18, 1229–1240. doi: 10.1016/S1473-3099(18)30475-4

Lamberti, L. M., Bourgeois, A. L., Fischer Walker, C. L., Black, R. E., and Sack, D. (2014). Estimating Diarrheal Illness and Deaths Attributable to *Shigella*e and Enterotoxigenic *Escherichia coli* among Older Children, Adolescents, and Adults in South Asia and Africa. *PLoS Negl. Trop. Dis.* 8:e2705. doi: 10.1371/journal.pntd.0002705

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Y., Luo, Q., Shi, X., Lin, Y., Qiu, Y., Lv, D., et al. (2017). Phenotypic and genotypic characterization of clinical enterotoxigenic *Escherichia coli* isolates from Shenzhen, China. *Foodborne Pathog. Dis.* 14, 333–340. doi: 10.1089/fpd.2016.2233

Liu, J., Bai, L., Li, W., Han, H., Fu, P., Ma, X., et al. (2018). Trends of foodborne diseases in China: lessons from laboratory-based surveillance since 2011. *Front. Med.* 12:48–57. doi: 10.1007/s11684-017-0608-6

Luo, Q., Kumar, P., Vickers, T. J., Sheikh, A., Lewis, W. G., Rasko, D. A., et al. (2014). Enterotoxigenic *escherichia coli* secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. *Infect. Immun.* 82, 509–521. doi: 10.1128/IAI.01106-13

Margulieux, K. R., Srijan, A., Ruekit, S., Nobthai, P., Poramathikul, K., Pandey, P., et al. (2018). Extended-spectrum β-lactamase prevalence and virulence factor characterization of enterotoxigenic *Escherichia coli* responsible for acute diarrhea in Nepal from 2001 to 2016. *Antimicrob. Resist. Infect. Control* 7, 1–7. doi: 10.1186/s13756-018-0377-2

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Nada, R. A., Shaheen, H. I., Khalil, S. B., Mansour, A., El-Sayed, N., Touni, I., et al. (2011). Discovery and phylogenetic analysis of novel members of class b enterotoxigenic *Escherichia coli* adhesive fimbriae. *J. Clin. Microbiol.* 49, 1403–1410.

Oh, K.-H., Kim, D. W., Jung, S.-M., and Cho, S.-H. (2014). Molecular Characterization of Enterotoxigenic *Escherichia coli* Strains Isolated from Diarrheal Patients in Korea during 2003–2011. *PLoS One* 9:e96896. doi: 10.1371/journal.pone.0096896

Pan, H., Zhang, J., Kuang, D., Yang, X., Ju, W., Huang, Z., et al. (2015). Molecular analysis and antimicrobial susceptibility of enterotoxigenic *Escherichia coli* from diarrheal patients. *Diagn. Microbiol. Infect. Dis.* 81, 126–131. doi: 10.1016/j.diagmicrobio.2014.10.008

Pilonieta, M. C., Bodero, M. D., and Munson, G. P. (2007). CfaD-Dependent Expression of a Novel Extracytoplasmic Protein from Enterotoxigenic *Escherichia coli*. *J. Bacteriol.* 189, 5060–5067. doi: 10.1128/JB.00131-07

Qadri, F., Svennerholm, A.-M., Faruque, A. S. G., and Sack, R. B. (2005). Enterotoxigenic *Escherichia coli* in Developing Countries: epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clin. Microbiol. Rev.* 18, 465–483. doi: 10.1128/CMR.18.3.465-483.2005

Qu, M., Zhang, X., Qian, H., Lyu, B., Huang, Y., Yan, H., et al. (2014). Study on the genotype and epidemic characteristics of diarrheagenic *Escherichia coli* isolated from diarrheal patients in Beijing]. *Zhonghua Liu Xing Bing Xue Za Zhi* 35, 1123–1126.

Rasko, D. A., Del Canto, F., Luo, Q., Fleckenstein, J. M., Vidal, R., and Hazen, T. H. (2019). Comparative genomic analysis and molecular examination of the diversity of enterotoxigenic *Escherichia coli* isolates

from Chile. *PLoS Negl. Trop. Dis.* 13:e0007828. doi: 10.1371/journal.pntd.000 7828

Rivera, F. P., Ochoa, T. J., Maves, R. C., Bernal, M., Medina, A. M., Meza, R., et al. (2010). Genotypic and Phenotypic Characterization of Enterotoxigenic *Escherichia coli* Strains Isolated from Peruvian Children. *J. Clin. Microbiol.* 48, 3198–3203. doi: 10.1128/JCM.00644-10

Ronholm, J., Nasheri, N., Petronella, N., and Pagotto, F. (2016). Navigating microbiological food safety in the era of whole-genome sequencing. *Clin. Microbiol. Rev.* 29, 837–857. doi: 10.1128/CMR.00056-16

Sahl, J. W., Sistrunk, J. R., Baby, N. I., Begum, Y., Luo, Q., Sheikh, A., et al. (2017). Insights into enterotoxigenic *Escherichia coli* diversity in Bangladesh utilizing genomic epidemiology. *Sci. Rep.* 7:3402. doi: 10.1038/s41598-017-03631-x

Sahl, J. W., Steinsland, H., Redman, J. C., Angiuoli, S. V., Nataro, J. P., Sommerfelt, H., et al. (2011). A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect. Immun.* 79, 950–960. doi: 10.1128/IAI.00932-10

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

Sjöling, Å, von Mentzer, A., and Svennerholm, A. M. (2015). Implications of enterotoxigenic *Escherichia coli* genomics for vaccine development. *Expert Rev. Vaccines* 14, 551–560. doi: 10.1586/14760584.2015.996553

So, M., Boyer, H. W., Betlach, M., and Falkow, S. (1976). Molecular cloning of an *Escherichia coli* plasmid determinant than encodes for the production of heat-stable enterotoxin. *J. Bacteriol.* 128, 463–472.

So, M., Dallas, W. S., and Falkow, S. (1978). Characterization of an *Escherichia coli* plasmid encoding for synthesis of heat-labile toxin: molecular cloning of the toxin determinant. *Infect. Immun.* 21, 405–411.

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

Torres, O. R., González, W., Lemus, O., Pratdesaba, R. A., Matute, J. A., Wiklund, G., et al. (2015). Toxins and virulence factors of enterotoxigenic *Escherichia coli* associated with strains isolated from indigenous children and international visitors to a rural community in Guatemala. *Epidemiol. Infect.* 143, 1662–1671. doi: 10.1017/S0950268814002295

Turner, S. M., Chaudhuri, R. R., Jiang, Z. D., DuPont, H., Gyles, C., Penn, C. W., et al. (2006). Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. *J. Clin. Microbiol.* 44, 4528–4536. doi: 10.1128/JCM.01474-06

Vidal, R. M., Muhsen, K., Tennant, S. M., Svennerholm, A.-M., Sow, S. O., Sur, D., et al. (2019). Colonization factors among enterotoxigenic *Escherichia coli* isolates from children with moderate-to-severe diarrhea and from matched

controls in the Global Enteric Multicenter Study (GEMS). *PLoS Negl. Trop. Dis.* 13:e0007037. doi: 10.1371/journal.pntd.0007037

von Mentzer, A., Blackwell, G. A., Pickard, D., Boinett, C. J., Joffré, E., Page, A. J., et al. (2021). Long-read-sequenced reference genomes of the seven major lineages of enterotoxigenic *Escherichia coli* (ETEC) circulating in modern time. *Sci. Rep.* 11:9256. doi: 10.1038/s41598-021-88316-2

von Mentzer, A., Connor, T. R., Wieler, L. H., Semmler, T., Iguchi, A., Thomson, N. R., et al. (2014). Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat. Genet.* 46, 1321–1326. doi: 10. 1038/ng.3145

von Mentzer, A., Tobias, J., Wiklund, G., Nordqvist, S., Aslett, M., Dougan, G., et al. (2017). Identification and characterization of the novel colonization factor CS30 based on whole genome sequencing in enterotoxigenic *Escherichia coli* (ETEC). *Sci. Rep.* 7:12514. doi: 10.1038/s41598-017-12743-3

Wolf, M. K. (1997). Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*. *Clin. Microbiol. Rev.* 10, 569–584.

Xiang, Y., Wu, F., Chai, Y., Xu, X., Yang, L., Tian, S., et al. (2020). A new plasmid carrying mphA causes prevalence of azithromycin resistance in enterotoxigenic *Escherichia coli* serogroup O6. *BMC Microbiol.* 20:247. doi: 10.1186/s12866-020-01927-z

Yang, C., Pei, X., Wu, Y., Yan, L., Yan, Y., Song, Y., et al. (2019). Recent mixing of Vibrio parahaemolyticus populations. *ISME J.* 13, 2578–2588. doi: 10.1038/ s41396-019-0461-5