Phylogenetic Analysis Reveals Distinct Evolutionary Trajectories of the Fluoroquinolones-Resistant *Escherichia coli* ST1193 From Fuzhou, China

Jiangqing Huang[1], Zhichang Zhao[2], Qianwen Zhang[1], Shengcen Zhang[1], Shuyu Zhang[3], Min Chen[3], Hongqiang Qiu[2], Yingping Cao[1]* and Bin Li[1]*

[1] Department of Clinical Laboratory, Fujian Medical University Union Hospital, Fuzhou, China, [2] Department of Pharmacy, Fujian Medical University Union Hospital, Fuzhou, China, [3] Department of Laboratory Medicine, Fujian Medical University, Fuzhou, China

Escherichia coli (*E. coli*) ST1193 is an emerging fluoroquinolones-resistant and virulent lineage. Large gaps remain in our understanding of the evolutionary processes and differences of this lineage. Therefore, we used 76 *E. coli* ST1193 genomes to detect strain-level genetic diversity and phylogeny of this lineage globally. All *E. coli* ST1193 possessed *fimH64*, *filCH5,* and *fumC14*. There was 94.7% of isolates classified as O-type O75. There was 9.33% of *E. coli* ST1193 that possessed K5 capsular, while 90.67% of isolates possessed K1 capsular. The core genome analysis revealed that all isolates were divided into two phylogenetic clades (clade A and B). Clade A included 25 non-Chinese *E. coli* ST1193, and clade B contained all isolates collected from Fuzhou, China, respectively. The results of comparative genomics indicated Indels were identified in 150 clade-specific genes, which were enriched into the biological process and molecular function. Accessory genome phylogenetic tree showed a high degree of correlation between accessory genome clusters and core genome clades. There was significant difference in antibiotic resistance genes (ARGs) [$bla_{CTX-M-55}$, bla_{TEM-1}, *sul2*, *tet(B)*, *tet(R)*, *APH(6)-Id,* and *AAC(3)-IId*], virulence factors (*cia*, *neuC*, *gad,* and *traT),* and plasmid replicon types (IncQ1, Col156, and IncB/O/K/Z) between clade A (non-Chinese isolates) and clade B (Chinese isolates) (*p* < 0.05). Further analysis of the genetic environments of $bla_{CTX-M-55}$ demonstrated that the flanking contexts of $bla_{CTX-M-55}$ were diverse. In conclusion, our results reveal the distinct evolutionary trajectories of the spread of *E. coli* ST1193 in Fuzhou, China and non-China regions. This supports both global transmission and localized lineage expansion of this lineage following specific introductions into a geographic locality.

Keywords: *E. coli*, ST1193, whole genome sequencing, phylogenomic analysis, $bla_{CTX-M-55}$, fluoroquinolones-resistant

INTRODUCTION

Escherichia coli (*E. coli*) is the most common gram-negative bacterial opportunistic pathogen. It primarily inhabits the lower intestinal tract of humans and other warm-blooded animals, presenting both a clinical and an epidemiological challenge (Vila et al., 2016). *E. coli* could cause many clinical infections, ranging from diarrhea, uncomplicated urinary tract infections to life-threatening septicemia, leading to more than two million deaths each year worldwide (Gomes et al., 2016; Tandogdu and Wagenlehner, 2016; Micenková et al., 2017; Askari and Ghanbarpour, 2019).

Multi-locus sequence typing (MLST) is a convenient sequence-based method for identifying clonal relationships among bacteria and is usually applied to detect the population structure of *E. coli* strains with respect to their specificity for infecting humans or animals (Chaudhuri and Henderson, 2012; Riley, 2014). In the past 20 years, resistance to fluoroquinolones and extended-spectrum cephalosporins rapidly increased due to the production of extended-spectrumβ-lactamases by *E. coli* (Nicolas-Chanoine et al., 2014). This resistance increase is linked to the worldwide spread of a specific clone of *E. coli*, sequence type (ST) 131 (Nicolas-Chanoine et al., 2014). Previous studies demonstrated that *E. coli* ST131 accounted for nearly half in fluoroquinolones-resistant *E. coli* (Tchesnokova V. et al., 2019; Tchesnokova V. L. et al., 2019). Interestingly, the *E. coli* population structure is dynamic (Johnson T. J. et al., 2019). A new sequence type, *E. coli* ST1193, was reported as an emerging fluoroquinolones-resistant (FQ-R) and virulent *E. coli* lineage that usually causes various extraintestinal infections, such as septicemia, urinary tract infection, and meningitis (Wu et al., 2017; Johnson J. R. et al., 2019; Ding et al., 2021). This lineage was usually isolated from animals and humans in many countries, including the United States, China, and Australia (Platell et al., 2012; Wu et al., 2017; Xia et al., 2017; Johnson T. J. et al., 2019). Moreover, the percentage of this lineage rose dramatically from 4.4 to 22.2% during 2010–2017 and the growing frequency of this linage soared fivefold over the ST131 linage (Tchesnokova V. et al., 2019). Until now, *E. coli* ST1193 became a sizable second clonal group behind ST131-H30. Therefore, the increase and rapid dissemination of *E. coli* ST1193 are of great concern.

Most of the studies recently focused on the epidemiological investigation among *E. coli* ST1193. A previous study illustrated probable key events leading to *E. coli* ST1193 lineage, but it could not verify geographical isolation sources and gain a solid conclusion regarding geographical clustering (Johnson T. J. et al., 2019). Information on geographical clustering in Chinese ST1193 strains is limited. Large gaps remain in our understanding regarding the evolutionary processes and differences of this lineage. This study aimed to detect strain-level genetic diversity and phylogeny of *E. coli* ST1193 lineage using whole genome sequencing and gain a deep insight into this lineage globally, and to investigate the presence of genetic elements related to antimicrobial resistance (AMR) that might confer selective advantages to this emerging bacterial pathogen.

MATERIALS AND METHODS

Bacterial Isolates and Whole Genome Sequencing

A total of 76 *E. coli* ST1193 were used in this study. Among them, 51 isolates were collected at Fujian Medical University Union Hospital (Fuzhou, China) between August 2014 and August 2015, reported in our previous study (Wu et al., 2017). These 51 *E. coli* ST1193 isolates were subjected to whole genome sequencing (WGS). Isolates were grown overnight from a single colony in 2 ml LB broth. The genomic DNA samples were extracted from *E. coli* ST1193 isolates by TIANamp Bacteria DNA Kit (Tiangen, Beijing) according to the manufacturer's instructions. They were sequenced using the Illumina NovaSeq platform at Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China). Then, FastQC software (version.0.11.7)[1] was used to detect the quality of raw data and data assembly was proceeding after adapter contamination removing and data filtering using AdapterRemoval (version.2.2.2) (Lindgreen, 2012) and SOAPec (version.2.03) (Luo et al., 2012) based on Kmer frequency. The filtered reads were assembled by SPAdes (version. 3.12.0) (Bankevich et al., 2012), and A5-miseq (version 20160825) (Coil et al., 2015) to constructed scaffolds and contigs. Pilon software (version 1.18) (Walker et al., 2014) was used for base correction. Genome annotation for each assembly was performed using Prokka (version.1.12) (Seemann, 2014) with *E. coli* serving as the target species. Those 51 *E. coli* ST1193 genomes were submitted to GenBank under Bioproject PRJNA670319 and released on January 2, 2021.

Enterobase was a repository for *E. coli* genomic data[2], and was also searched for isolates belonging to ST1193 (according to the Achtman 7-gene MLST scheme, including *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*) with accession numbers linked to the National Center for Biotechnology Information (NCBI) database[3] (Zhou et al., 2020). The remaining 25 *E. coli* ST1193 were obtained from NCBI. A total of 25 assembled *E. coli* ST1193 genomes were released on NCBI on April 1, 2020, which were included in this study for further analysis. The Accession numbers for these 25 isolates are listed in **Supplementary Table 1**. Moreover, the metadata information (strain name, source, isolated date, and country of origin) was collected and coupled with the phylogenetic tree for further analysis. All 25 *E. coli* ST1193 genomes were provided with new annotation using Prokka.

Genome Data Analysis

The Center for Genomic Epidemiology tools[4] were used to identify the gene contents. Specifically, plasmid types were identified using PlasmidFinder (version 2.0.1) and pMLST (version 0.1.0) (95% minimum match and 60% minimum length); virulence factors (VFs) were detected using VirulenceFinder (version 2.0.3) (90% minimum match and 60% minimum length);

---

[1]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[2]http://enterobase.warwick.ac.uk/
[3]https://www.ncbi.nlm.nih.gov/
[4]https://cge.cbs.dtu.dk/services/

the multilocus-sequence typing was analyzed using MLST (version 2.0.4); serotypes and *fimH* alleles were predicted with SerotypeFinder (version 2.0.1) (85% minimum match and 60% minimum length) and FimTyper (version 1.0) (95% minimum match), respectively.

In addition, CARD (version 3.1.3)[5] was used to determine resistance genes and known chromosomal mutations conferring antibiotic resistance (Alcock et al., 2020). CRISPRcasFinder[6] (Couvin et al., 2018) was used with default parameters to identify the CRISPR loci in the genomes and determine the number and sequences of the spacers within CRISPR repeat arrays. The only CRISPR with evidence-levels 3 and 4 were considered as highly likely CRISPR arrays in this study according to the instructions (Couvin et al., 2018).

Core Genome Analysis and Accessory Genome Analysis

First, the sequence was classified into gene families. The potential homologous genes were detected by sequence similarity and all homologous genes in each *E. coli* ST1193 genome were clustered into a single gene family by OrthoFinder software (version 2.3.12) (Emms and Kelly, 2015). The BlastP tool (Li and Lu, 2019) was used to compare the protein sequences from all genomes with an e-value cutoff set at 10^{-10} and an additional matching length criterion set at 85% of the query sequence. In order to further analyze the gene family, the mafft (version 7.429) software (Katoh et al., 2002) was applied to compare the sequence of gene families. Meanwhile, the sequences were annotated using clusters of orthologous groups of proteins (COGs) database by BlastP with an e-value cutoff of 10^{-10} as described previously (Li et al., 2017).

Core genomes were defined as described previously (McNally et al., 2016). Based on the comparative analysis of 76 *E. coli* ST1193, core genomes were selected as single-copy core-genome for sequence alignment using mafft (Cazares et al., 2020). The set of single-copy core-genome clusters detected was entered into the pipeline to identify high-quality phylogenetic markers and infer a core genome phylogeny through a maximum-likelihood tree search. Following extraction of core genome sequences, the remaining sequences of 76 *E. coli* ST1193 were designated as accessory genomes (Ozer et al., 2014). The data set of accessory genomes was composed of single-copy genes shared by 50% or more genomes, but less than 100% genomes and those accessory genomes were selected for sequence alignment using mafft. Maximum likelihood phylogeny of core and accessory genome of 76 *E. coli* were inferred from the alignment using FastTree (version 2.1.11) with Whelan-and-Goldman 2001 model and 1,000 bootstrap replicates, respectively (Price et al., 2009). The resulting phylogenetic tree was visualized using iTOL (version 6) which was also used to overlay metadata information (Letunic and Bork, 2019). Furthermore, the results of clusters of accessory genome phylogenetic tree were visualized in combination with the corresponding core genome phylogenetic tree using iTOL.

Gene Ontology Term Enrichment Analysis

The SNP and indels of differentiated genes of clades in the core genome phylogenetic tree were identified using mafft (version 7.429). Then, those differentiated genes were defined as clade-specific genes and submitted to gene ontology (GO) term enrichment analysis by GOEAST (version 1.2.0)[7].

Statistical Analysis

Statistical analysis was performed using SAS Statistical Software (version 9.4). The Chi-square test (χ^2) or Fisher's exact test (two-tailed) was performed to compare data [including virulence factors, plasmids replicon types, and antibiotic resistance genes (ARGs)] of two groups. For each comparison, only a *p*-value < 0.05 was statistically significant.

RESULTS

Molecular Characteristics of *Escherichia coli* ST1193

A total of 76 *E. coli* ST1193 genome sequences were included in this study. All of them were collected from humans from Fuzhou, China and non-China regions with a broad range of years. All *E. coli* ST1193 possessed *fimH64*, *filCH5*, and *fumC14* allele. There was 94.7% (72/76) of *E. coli* ST1193 classified as O-type O75. The remaining four isolates could not be typed.

Two types of capsule variants (K1 and K5) were found in this study. Among them, seven (9.33%) *E. coli* ST1193 possessed K5 capsular genotype, while 68 (90.67%) isolates possessed K1 capsular genotype. One isolate possessed neither K1 capsular genotype nor K5 capsular genotype. There was 16.67% (4/24) of non-Chinese *E. coli* ST1193 that possessed K5 capsular and the percentage was higher than that in isolates collected from Fuzhou, China (5.88%, 3/51).

The highly likely CRISPR arrays were not detected in all of *E. coli* ST1193 in this study. The results of *fimH* and *fumC* alleles, capsule variants, and O-types in *E. coli* ST1193 and the metadata information of each isolate are shown in **Figure 1**.

Phylogenomic Analysis of *Escherichia coli* ST1193

In total, 3,454 single-copy core-genes were detected in the alignment. The core-genome alignments were applied for phylogenetic tree reconstruction using maximum likelihood estimation (**Figure 1**). All *E. coli* ST1193 were divided into two phylogenetic clades (clade A and B), with each supported by 100% bootstrapping. Clade A included 25 non-Chinese *E. coli* ST1193, and clade B contained all isolates from Fuzhou, China, respectively. Meanwhile, the isolates from clade B were more diverse than the entire worldwide collection. We further explored the difference of distinct clades in the core genome phylogenetic tree. The results of comparative genomics indicated that Indels were identified in a total of 150 clade-specific genes in our
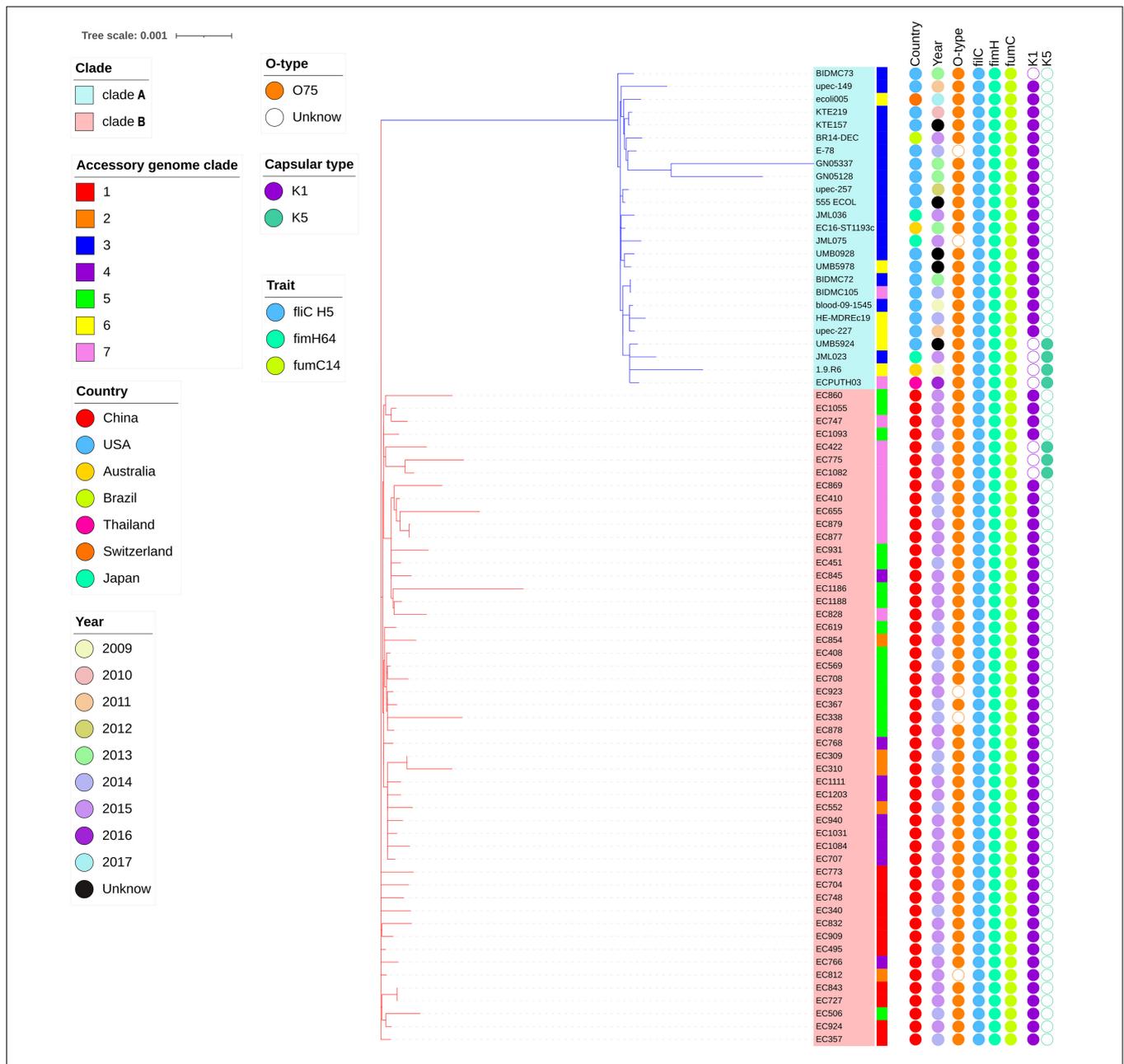
FIGURE 1 | Phylogenetic tree of *E. coli* ST1193 based on single-copy gene in core genome. The phylogram was built from 3,454 single-copy core-genes using maximum likelihood (ML). Branch support was performed by Whelan-and-Goldman 2001 model and 1,000 bootstrap replicates. Taxon labels for clades A and B are colored blue and red, respectively. The accessory genome clades information is shown as colored strips surrounding the phylogram. Metadata are represented as circles as follows: year and geographical region in assorted colors as depicted in the legend. Allelic profiling information is shown as colored circles surrounding the phylogram (from left to right) for the O-type, *fliC*, *fimH*, *fumC*, K1, and K5 capsular genes.

study (**Supplementary Table 2**). Based on the presence of these indels in particular genes, 76 *E. coli* ST1193 could be clearly clustered into two clades by geographical distribution. Besides, GO term enrichment analysis results showed that more than 70% of genes were enriched into the biological process and molecular function (**Figure 2**).

Furthermore, a total of 1,134 accessory genes were extracted to create an accessory genome matrix for all 76 *E. coli*

ST1193 genomes. The accessory genome phylogenetic tree was obtained using maximum likelihood estimation (**Supplementary Figure 1**), resulting in seven distinct clades (clade 1–7). We mapped accessory genome clusters onto the core genome phylogeny (**Figure 1**). The results showed a high degree of correlation between accessory genome clusters (clades 3 and 6) and core genome phylogeny in clade A. Meanwhile, multiple accessory genome clusters were distributed throughout core

FIGURE 2 | GO enrichment analysis result of 150 clade-specific genes.
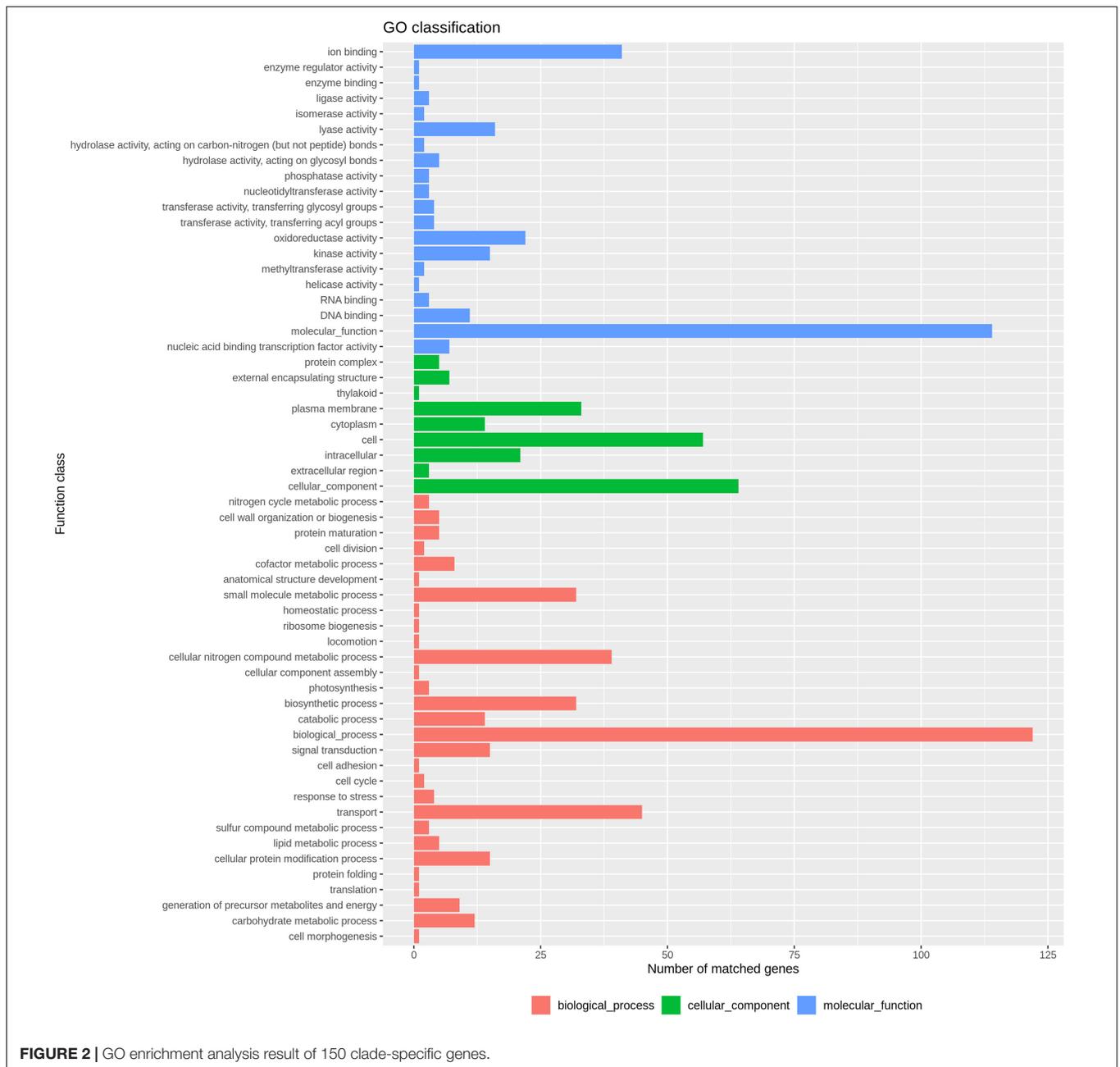
genome phylogeny clade B that revealed the 51 *E. coli* ST1193 from Fuzhou, China with high genetic diversity.

Virulence Factors

The distribution of VFs differed among isolates (**Figure 3**). Some VFs were identified more than 70% among *E. coli* ST1193, such as outer membrane hemin receptor gene *chuA* (76/76, 100%), siderophore receptor gene *fyuA* (76/76, 100%), fimbrial protein gene *yfcV* (76/76, 100%), vacuolating autotransporter gene *vat* (76/76, 100%), asecreted autotransporter gene *sat* (72/76, 94.7%), IrgA homolog adhesin gene *iha* (71/76, 93.4%), glutamate decarboxylase gene *gad* (70/76, 92.1%), and plasmid-carried

enterotoxin gene *senB* (58/76, 76.3%). These VFs mentioned above could be detected across all geographical regions.

Notably, there were significant differences in VFs between clade A (non-Chinese *E. coli* ST1193 isolates) and clade B (isolates from Fuzhou, China), including *cia*, *neuC*, *gad*, and *traT* (**Figure 4A**, $p < 0.05$). However, those differences were slight compared to the entire VFs profiles.

Plasmids Replicon Types

Overall, we found that almost all isolates (75/76, 98.7%) carried at least one plasmid replicon type, covering 19 different known plasmid incompatibility types. Only one isolate (1.3%) contained
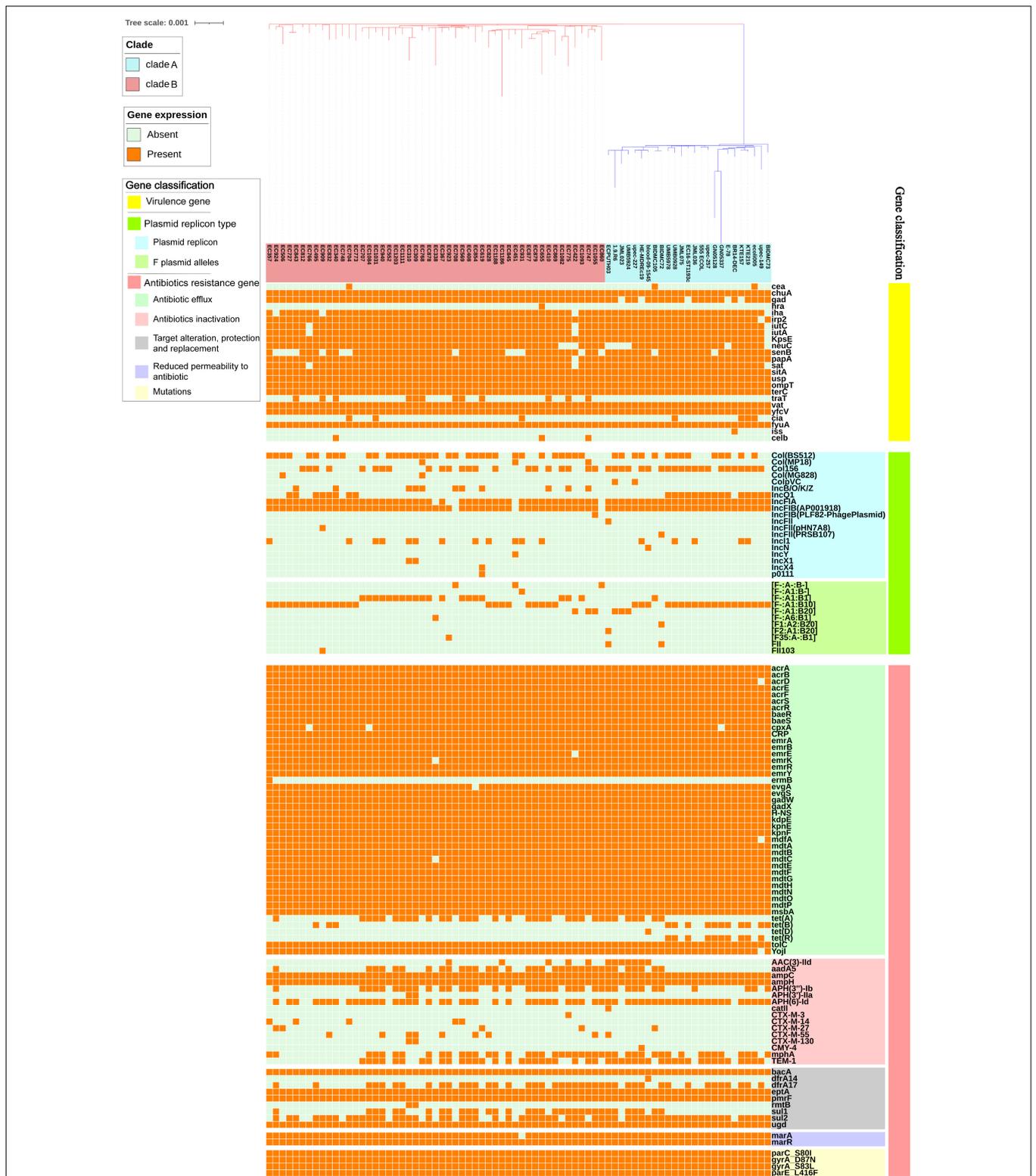
FIGURE 3 | The virulence factors, plasmid replicon types, and antibiotics resistance genes in *E. coli* ST1193 genome. Heat map of 76 *E. coli* ST1193 genomes (51 isolates from Fuzhou, China and 25 isolates from non-China regions) showing the presence and absence of 92 antibiotic resistance genes (belonging to different categories, including antibiotic efflux, antibiotic inactivation, target alteration, protection and replacement, and reduced permeability to antibiotics) and four mutations in housekeeping genes; 24 virulence factors; 19 plasmid type and 11 F plasmid alleles. Gene names are represented on the Y axis and the *E. coli* ST1193 genomes are listed on the X axis.

FIGURE 4 | The comparison of virulence factors **(A)**, plasmid replicon types **(B)**, and antibiotics resistance genes **(C)** between isolates from non-China region (clade A) and isolates from Fuzhou, China (clade B). The comparison data were categorical variables and the Chi-square test (χ^2) or Fisher's exact test (two-tailed) were performed for data comparison of two groups. Only $p < 0.05$ was statistically significant. *$p < 0.05$.

no plasmid incompatibility types (**Figure 3**). Most *E. coli* ST1193 (73/76, 96.1%) contained two types of F-type replicons (IncFIA and IncFIB), while IncFII replicon was only detected in three strains. Plasmid multilocus sequence typing (MLST) based on F plasmid alleles was performed in this study. Our results revealed that IncFIA allele (A1) was highly prevalent across *E. coli* ST1193.

IncFIB allele was diverse and four different alleles (B-, B1, B10 and B20) of IncFIB were identified. Col (BS512) and Col156 were highly prevalent among isolates from all geographical regions in this study.

Additionally, there were significant differences in plasmids replicon types carried between clade A (non-Chinese *E. coli* ST1193 isolates) and clade B (isolates from Fuzhou, China) (**Figure 4B**). IncQ1 was identified in 60.9% of non-Chinese *E. coli* ST1193, but found in only 39.1% of isolates from Fuzhou, China ($p < 0.05$). Col156 was detected in 84.0% of non-Chinese *E. coli* ST1193, but found in only 43.2% of isolates from Fuzhou, China ($p < 0.05$). IncB/O/K/Z was commonly found among *E. coli* ST1193 from Fuzhou, China (11/51, 21.6%), and undetected in non-Chinese isolates (0/25, 0%) ($p < 0.05$).

Antibiotic Resistance Genes

The distribution of ARGs among *E. coli* ST1193 is shown in **Figure 3**. In this study, all ST1193 harbored the same four non-synonymous mutations in *parC* (S80I), *parE* (L416F), and *gyrA* (D87N and S83L) housekeeping genes (**Figure 3**). Mutations in those four housekeeping genes are known to confer fluoroquinolone resistance. Some antibiotic efflux genes and genes associated with reduced permeability to antibiotics were present in all 76 *E. coli* ST1193, such as *acrA*, *acrB*, *emrE*, and *mdtA*. Among *E. coli* ST1193, more than 50% isolates carried acquired antimicrobial resistance genes, including bla_{TEM-1B} (ampicillin resistance, 60.5%), *sul2* (sulfonamide resistance, 77.6%), *APH(6)-Id* (aminoglycoside resistance, 76.4%), *dfrA17* (trimethoprim resistance, 55.3%), and *mphA* (macrolide resistance, 57.9%). Notably, $bla_{CTX-M-130}$ and $bla_{CTX-M-14}$ were only detected in clade A (non-Chinese *E. coli* ST1193 isolates) and *tet(B)* and *tet(R)* were only identified in clade B (isolates from Fuzhou, China), respectively.

Furthermore, we observed significant differences in ARGs carried between clade A and clade B (**Figure 4C**, $p < 0.05$), including $bla_{CTX-M-55}$, bla_{TEM-1}, *sul2*, *tet(B)*, *tet(R)*, *APH(6)-Id*, and *AAC(3)-IId*.

$bla_{CTX-M-55}$ With Different Genetic Environments in *Escherichia coli* ST1193

$bla_{CTX-M-55}$ was usually found in *E. coli* ST1193 as described previously and might play an important role in clonal expansion of ST1193 lineage, like $bla_{CTX-M-15}$ in ST131 (Stoesser et al., 2016; Xia et al., 2017). This prompted us to further examine the genetic environment of this gene in ST1193. In our study, $bla_{CTX-M-55}$ was found in nine isolates which all collected from Asia, including eight from Fuzhou, China and one from Thailand (**Figure 5**). From $bla_{CTX-M-55}$ gene were present in plasmid-associated sequences in the contig assemblies, we could determine precise plasmid–AMR gene links for those isolates. In addition, we found that $bla_{CTX-M-55}$ and IncI1 and/or FIB were identified in the same contig in 55.6% of isolates. In all nine isolates, $bla_{CTX-M-55}$ was located on the downstream of a homologous tract of ORF477. Different mobile elements were detected on the downstream or the upstream of $bla_{CTX-M-55}$, including *ISEc9*, *Tn3*, and integrase gene.

bla_{TEM-1} is located on the upstream of $bla_{CTX-M-55}$ in two isolates from Fuzhou, China.

DISCUSSION

Escherichia coli ST1193 is a global pandemic strain and increasingly linked to various infections (Tchesnokova V. L. et al., 2019). The emergence and clinical significance of *E. coli* ST1193 suggested that fluoroquinolones-resistant is the signature feature of pandemic ST1193 clonal group (Wu et al., 2017; Tchesnokova V. et al., 2019). *E. coli* ST1193 spread rapidly worldwide during the last years (Tchesnokova V. L. et al., 2019). The increase and dissemination of *E. coli* ST1193 are of great concern. Although there has been increased research on *E. coli* ST1193 epidemiology and phylogenomics, the development of multidrug resistance and genomic diversification of ST1193 is not yet clear. In this study, we reported the whole genome sequences of 76 *E. coli* ST1193 from different countries. Combining analysis of core genome and accessory genome phylogeny, virulence factors, antibiotics resistance genes, and plasmid replicon profiles, we provide a significant insight for the important evolutionary characteristics and differences of *E. coli* ST1193, which supported its emergence as a dominant and successful MDR pathogen.

Based on genomic observations, we highlighted key characteristics of *E. coli* ST1193. Most of these strains were serotype O75. They all possessed *fimH*64, *filC*14, and *fumC*14 allele, which were similar to a previous study (Johnson T. J. et al., 2019). Furthermore, the CRISPR arrays were not found in any *E. coli* ST1193 in this study which was similar to a previous study (Al-Farsi et al., 2020).

WGS is an effective way to analyze the evolution and spread of bacterial strains (Knight et al., 2015; Meehan et al., 2019). Here, we combined core and accessory genomes to explore evolutionary patterns of separation within the population. A previous study demonstrated that *E. coli* ST1193 could be clustered to many distinct clades, but it could not verify geographical isolation sources and gain a confident conclusion regarding geographical clustering (Johnson T. J. et al., 2019). However, our study provides a new insight into the phylogeny of *E. coli* ST1193. There were 76 strains that could be clustered into two clades based on origin (Fuzhou, China and non-China) in a core genome phylogeny tree (**Figure 1**). Clade A contained 25 non-Chinese *E. coli* ST1193, and clade B included 51 *E. coli* ST1193 collected from Fuzhou, China. The results of core genome phylogenomic analysis of *E. coli* ST1193 indicates large differences between isolates from Fuzhou, China and non-China region, suggesting complex evolutionary processes underlying ST1193. A comparative genomics analysis was performed to demonstrate the reasons for the differences between two distance clades. Notably, indels were identified in a total of 150 clade-specific genes that enriched the biological process and molecular function in this study (**Figure 2**). Conserved indels of defined length and sequence were found at the same position in a given protein (or gene) in all members from one or more groups are useful to understand the evolutionary relationship of bacteria (Gupta, 2003). Our results indicated 150 indels

FIGURE 5 | The genetic environments of *bla*$_{CTX-M-55}$ in *E. coli* ST1193. X, gene contexts limited by the extent of the assembled region around the *bla*$_{CTX-M-55}$. Red mark, *bla*$_{CTX-M-55}$ located on the same contig with plasmid replicon.
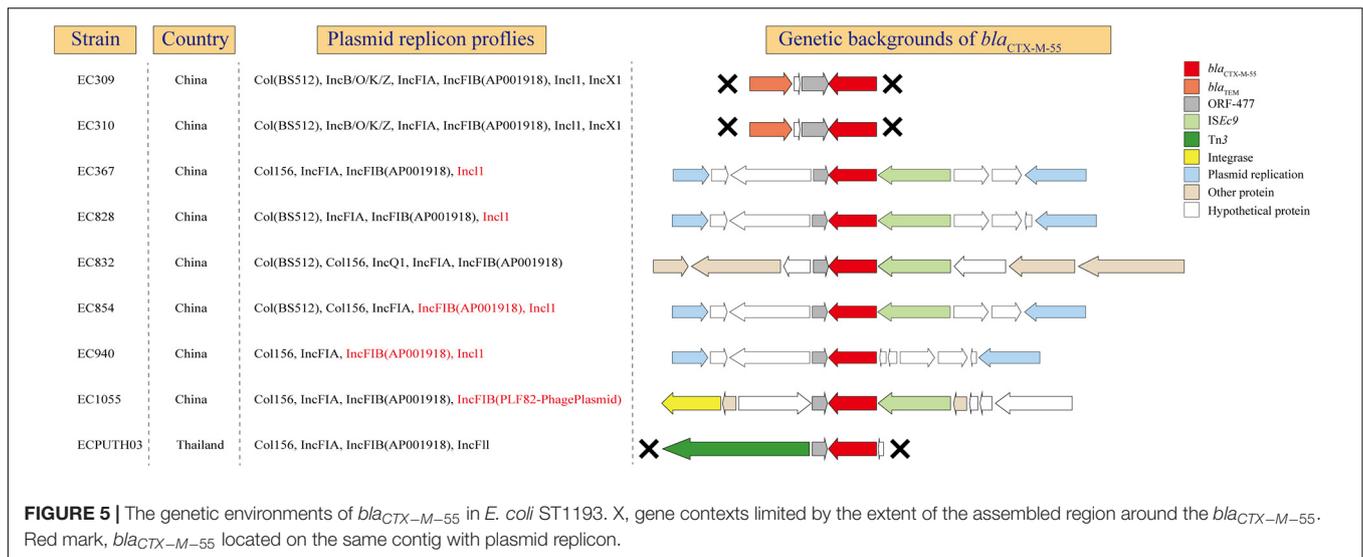
were responsible for regional differences in *E. coli* ST1193 and were contributed to determine the relative branching orders of different clades in clade A (harboring 25 non-Chinese *E. coli* ST1193) and clade B (harboring 51 *E. coli* ST1193 from Fuzhou, China). A recent study highlighted that it is possible to provide a completely different perspective of the evolution of a well-defined or undefined bacterial lineage by combining core and accessory genome analysis (McNally et al., 2016). The accessory genome clusters of bacteria were due to the unique combinations of genes circulating in the accessory gene pool (Reuter et al., 2015). Therefore, we mapped the accessory genome clusters onto core genome phylogeny that showed a high degree of correlation between accessory genome profile and core genome phylogeny in clade A, and then multiple accessory genome clusters distributed throughout clade B (**Figure 1** and **Supplementary Figure 1**). Our results indicated that *E. coli* ST1193 might contain multiple subtypes. Each of the subtypes arises due to the expansion of a successful clone following the emergence of a stable accessory gene profile. Interestingly, the results of core and accessory genome analysis demonstrated that geographical distributions might play a crucial role in the whole process throughout the phylogeny of *E. coli* ST1193. Furthermore, our study provided confident conclusions regarding geographical clustering for *E. coli* ST1193.

Escherichia coli ST1193 lineage exhibited a diversity of VFs profiles in this study. We found identical VFs profiles (*vat*, *fyuA*, *iutA*, *iha*, *gad*, *senB*, and *chuA*) among most ST1193 isolates (**Figure 3**). Those VFs were commonly prevalent in *E. coli* ST1193, consistent with a previous study (Valenza et al., 2019). Among those VFs, *senB* is a plasmid-carried enterotoxin gene (often detected in F-type plasmid) associated with enhancing bladder colonization and invasion (Cusumano et al., 2010). In addition, F-type plasmid replicons were most common in *E. coli* ST1193 in this study, including 100% in non-Chinese isolates and 94.2% in isolates from Fuzhou, China, respectively. Given this virulence factor's presence in many

successful ExPEC clones and the high prevalence of F-type plasmid in *E. coli* ST1193, it might be vital for ST1193's epidemiologic success (Cusumano et al., 2010). Furthermore, we found significant differences were found in VFs (*neuC*, *gad*, and *traT*) between two clades (**Figure 4A**, $p < 0.05$). Among them, *traT* was related to the serum resistance of *E. coli* (Daga et al., 2019). *neuC* was the prerequisite VF in bloodstream infections (Čurová et al., 2020). *gad* participated in acid resistance in *E. coli* (Giangrossi et al., 2005). Based on the above-mentioned evidence, we suspect that *E. coli* ST1193 from Fuzhou, China might process better survivability than non-Chinese isolates. Nonetheless, more experimental evidence is needed to support this conclusion because there is a similar abundance of other VFs between the two groups.

Escherichia coli ST1193 lineage exhibited a diversity of antibiotics resistance gene profiles. In this study, all *E. coli* ST1193 isolates harbored the same four non-synonymous mutations in the target genes—two in *gyrA* (D87N and S83L), one in *parC* (S80I), and one in *parE* (L416F) and those mutations confer fluoroquinolone resistance (**Figure 3**). Our results supported the notion that the mutant variants of *gyrA* and *parC* were acquired not by a typical gradual, stepwise evolution but all at once (Tchesnokova V. et al., 2019). Notably, *E. coli* ST1193 from Fuzhou, China and non-Chinese isolates contained different ESBLs genes, which might be related to the epidemiology of those ESBLs genes and medication habits of different regions. Interestingly, *tet(B)* and *tet(R)* showed a higher prevalence in non-Chinese isolates in our study ($p < 0.05$). *tet(B)* [or *tet(R)*] is a type of tetracycline resistance genes, usually found in *E. coli* isolated from animals and environment (Seyfried et al., 2010; Lee et al., 2021). The difference prevalence of *tet(B)* [or *tet(R)*] between *E. coli* ST1193 from Fuzhou, China and non-Chinese ST1193 indicated that there might be a probable foodborne pathway to transmit *tet(R)* [or *tet(B)*] to humans *via* the food chain in a specified area. Additionally, the highly prevalent IncQ1 plasmid was observed within non-Chinese *E. coli* ST1193

(clade A) ($p < 0.05$). IncQ1 plasmid played a crucial role in the spread of tetracycline resistance, which might explain the high prevalence of *tet* genes in this study (Sun et al., 2019; Dungan and Bjorneberg, 2020). Mobile genetic elements could mediate independent microevolution of bacteria, such as ARGs and plasmids (Wu et al., 2019). Therefore, high dissemination of *tet* gene and IncQ1 within non-Chinese *E. coli* ST1193 isolates might participate in the microevolution of this lineage in a specified area.

A previous study suggested that the wide distribution of bla_{CTX-M} variants is mediated by clonal expansion of CTX-M containing *E. coli* and its global dissemination (Stoesser et al., 2016). Our study showed the persistence of particular bla_{CTX-M} variants ($bla_{CTX-M-55}$) in *E. coli* ST1193 within specific geographic regions, such as Asia. $bla_{CTX-M-55}$ was highly prevalent among *E. coli* isolates in China (Xia et al., 2017). In addition, the spread of this gene in China was mainly attributed to the clonal dissemination of *E. coli* ST1193 (Xia et al., 2017). Thus, we further explored the potential mechanisms of this gene at genetic levels (gene, genetic environments, and plasmid) (**Figure 5**). Our results showed that $bla_{CTX-M-55}$ is located on IncI1 or IncFIB plasmids in most strains (55.6%), which was similar to a previous study from China (Xia et al., 2017). Meanwhile, the flanking contexts of $bla_{CTX-M-55}$ in those strains were diverse, indicating that some different mechanisms (such as horizontal gene transfer) might play a significant role in microevolution in ST1193 lineage. The diverse genetic environments of $bla_{CTX-M-55}$ might contribute to the global dissemination of this sequence type within specific geographic regions. As described previously, the almost ubiquitous presence of $bla_{CTX-M-15}$ within ST131 H30Rx is most striking and strongly associated with the presence of an IncFII_AY458016-like replicon (Stoesser et al., 2016). Therefore, further investigation should be performed to determine whether $bla_{CTX-M-55}$ within *E. coli* ST1193 and its associated replicon (IncI1 and/or IncFIB) contribute to subcluster and spread across geographic regions, like ST131 H30Rx.

There are several limitations to this study. First, since Chinese *E. coli* ST1193 were collected during a relatively short period of time, we did not elucidate the trends across years. Second, the short-read sequencing (Illumina NovaSeq platform) and limited transformant sequencing could not fully assess the flanking regions and plasmid structures across the entire data set. Third, the Chinese strains were only isolated from a single center (Fuzhou, China) and further study needs to increase the sample size, including new strains from other provinces of China. Meanwhile, the number of genomes from the rest of the world was very small. More strains from other countries should be included in further analyses to make the results of comparative genomics analysis more representatives in the future.

CONCLUSION

Our study provides an important insight into the phylogenetic relationship of fluoroquinolones-resistant *E. coli* ST1193. Our

results reveal that the distinct evolutionary trajectories of the spread of *E. coli* ST1193 in Fuzhou, China and non-China regions and the *E. coli* ST1193 isolates from Fuzhou, China seemed more diverse than the entire worldwide collection. That different phylogenetic relationship supports both global transmission and localized lineage expansion of this lineage following specific introductions into a geographic locality. Our results were a good pilotage analysis and could provide a start point for further extended analysis. Strategic and continued surveillance should be conducted to prevent the infections caused by *E. coli* ST1193.

DATA AVAILABILITY STATEMENT

The genome sequences of 51 *E. coli* ST1193 were deposited in GenBank under BioProject PRJNA670319 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA670319).

ETHICS STATEMENT

All procedures of this study involving clinical isolates and their associated metadata information were reviewed and approved by the Medical Ethics Committee of Fujian Medical University Union Hospital (2020KY0121).

AUTHOR CONTRIBUTIONS

JH, SCZ, SYZ, and QZ performed the experiments. JH drafted the manuscript. QZ and ZZ analyzed the data. YC, HQ, and MC supervised the study. BL designed the study and revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript and contributed to manuscript revision, read, and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.746995/full#supplementary-material

REFERENCES

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edaltmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Al-Farsi, H. M., Camporeale, A., Ininbergs, K., Al-Azri, S., Al-Muharrmi, Z., Al-Jardani, A., et al. (2020). Clinical and molecular characteristics of carbapenem non-susceptible *Escherichia coli*: a nationwide survey from Oman. *PLoS One* 15:e0239924. doi: 10.1371/journal.pone.0239924

Askari, N., and Ghanbarpour, R. (2019). Molecular investigation of the colicinogenic *Escherichia coli* strains that are capable of inhibiting E. coli O157:H7 *in vitro. BMC Vet. Res.* 15:14. doi: 10.1186/s12917-018-1771-y

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

Cazares, A., Moore, M. P., Hall, J. P. J., Wright, L. L., Grimes, M., Emond-Rhéault, J. G., et al. (2020). A megaplasmid family driving dissemination of multidrug resistance in *Pseudomonas. Nat. Commun.* 11:1370. doi: 10.1038/s41467-020-15081-7

Chaudhuri, R. R., and Henderson, I. R. (2012). The evolution of the *Escherichia coli* phylogeny. *Infect. Genet. Evol.* 12, 214–226. doi: 10.1016/j.meegid.2012.01.005

Coil, D., Jospin, G., and Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31, 587–589. doi: 10.1093/bioinformatics/btu661

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

Čurová, K., Slebodníková, R., Kmeťová, M., Hrabovský, V., Maruniak, M., Liptáková, E., et al. (2020). Virulence, phylogenetic background and antimicrobial resistance in *Escherichia coli* associated with extraintestinal infections. *J. Infect. Public Health* 13, 1537–1543. doi: 10.1016/j.jiph.2020.06.032

Cusumano, C. K., Hung, C. S., Chen, S. L., and Hultgren, S. J. (2010). Virulence plasmid harbored by uropathogenic *Escherichia coli* functions in acute stages of pathogenesis. *Infect Immun.* 78, 1457–1467. doi: 10.1128/iai.01260-09

Daga, A. P., Koga, V. L., Soncini, J. G. M., de Matos, C. M., Perugini, M. R. E., Pelisson, M., et al. (2019). *Escherichia coli* bloodstream infections in patients at a university hospital: virulence factors and clinical characteristics. *Front. Cell Infect. Microbiol.* 9:191. doi: 10.3389/fcimb.2019.00191

Ding, Y., Zhang, J., Yao, K., Gao, W., and Wang, Y. (2021). Molecular characteristics of the new emerging global clone ST1193 among clinical isolates of *Escherichia coli* from neonatal invasive infections in China. *Eur. J. Clin. Microbiol. Infect. Dis.* 40, 833–840. doi: 10.1007/s10096-020-04079-0

Dungan, R. S., and Bjorneberg, D. L. (2020). Antibiotic resistance genes, class 1 integrons, and IncP-1/IncQ-1 plasmids in irrigation return flows. *Environ. Pollut.* 257:113568. doi: 10.1016/j.envpol.2019.113568

Emms, D. M., and Kelly, S. (2015). Ortho finder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Giangrossi, M., Zattoni, S., Tramonti, A., De Biase, D., and Falconi, M. (2005). Antagonistic role of H-NS and GadX in the regulation of the glutamate decarboxylase-dependent acid resistance system in *Escherichia coli. J. Biol. Chem.* 280, 21498–21505. doi: 10.1074/jbc.M413255200

Gomes, T. A., Elias, W. P., Scaletsky, I. C., Guth, B. E., Rodrigues, J. F., Piazza, R. M., et al. (2016). Diarrheagenic *Escherichia coli. Braz J. Microbiol.* 47(Suppl. 1), 3–30. doi: 10.1016/j.bjm.2016.10.015

Gupta, R. S. (2003). Evolutionary relationships among photosynthetic bacteria. *Photosynth Res.* 76, 173–183. doi: 10.1023/a:1024999314839

Johnson, J. R., Johnston, B. D., Porter, S. B., Clabots, C., Bender, T. L., Thuras, P., et al. (2019). Rapid emergence, subsidence, and molecular detection of *Escherichia coli* sequence type 1193-fimH64, a new disseminated multidrug-resistant commensal and extraintestinal pathogen. *J. Clin. Microbiol.* 57:e1664-18. doi: 10.1128/jcm.01664-18

Johnson, T. J., Elnekave, E., Miller, E. A., Munoz-Aguayo, J., Flores Figueroa, C., Johnston, B., et al. (2019). Phylogenomic analysis of extraintestinal pathogenic *Escherichia coli* sequence type 1193, an emerging multidrug-resistant clonal group. *Antimicrob Agents Chemother.* 63:e1913-18. doi: 10.1128/aac.01913-18

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Knight, D. R., Elliott, B., Chang, B. J., Perkins, T. T., and Riley, T. V. (2015). Diversity and evolution in the genome of clostridium difficile. *Clin. Microbiol. Rev.* 28, 721–741. doi: 10.1128/cmr.00127-14

Lee, H. J., Woo, Y. K., Choi, B. K., Jeong, O. M., Kim, J. H., Kim, D. W., et al. (2021). High prevalence of a gene cluster conferring resistance to streptomycin, sulfonamide, and tetracycline in *Escherichia coli* isolated from indigenous wild birds. *J. Gen. Appl. Microbiol.* 67, 81–84. doi: 10.2323/jgam.2020.06.003

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

Li, N., Wang, K., Williams, H. N., Sun, J., Ding, C., Leng, X., et al. (2017). Analysis of gene gain and loss in the evolution of predatory bacteria. *Gene* 598, 63–70. doi: 10.1016/j.gene.2016.10.039

Li, Y. C., and Lu, Y. C. (2019). BLASTP-ACC: parallel architecture and hardware accelerator design for BLAST-based protein sequence alignment. *IEEE Trans. Biomed. Circ. Syst.* 13, 1771–1782. doi: 10.1109/tbcas.2019.2943539

Lindgreen, S. (2012). Adapter removal: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5:337. doi: 10.1186/1756-0500-5-337

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. doi: 10.1186/2047-217x-1-18

McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., et al. (2016). Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLoS Genet* 12:e1006280. doi: 10.1371/journal.pgen.1006280

Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., et al. (2019). Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat. Rev. Microbiol.* 17, 533–545. doi: 10.1038/s41579-019-0214-5

Micenková, L., Beňová, A., Frankovičová, L., Bosák, J., Vrba, M., Ševčíková, A., et al. (2017). Human *Escherichia coli* isolates from hemocultures: septicemia linked to urogenital tract infections is caused by isolates harboring more virulence genes than bacteraemia linked to other conditions. *Int. J. Med. Microbiol.* 307, 182–189. doi: 10.1016/j.ijmm.2017.02.003

Nicolas-Chanoine, M. H., Bertrand, X., and Madec, J. Y. (2014). *Escherichia coli* ST131, an intriguing clonal group. *Clin. Microbiol. Rev.* 27, 543–574. doi: 10.1128/CMR.00125-13

Ozer, E. A., Allen, J. P., and Hauser, A. R. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools spine and AGEnt. *BMC Genom.* 15:737. doi: 10.1186/1471-2164-15-737

Platell, J. L., Trott, D. J., Johnson, J. R., Heisig, P., Heisig, A., Clabots, C. R., et al. (2012). Prominence of an O75 clonal group (clonal complex 14) among non-ST131 fluoroquinolone-resistant *Escherichia coli* causing extraintestinal infections in humans and dogs in Australia. *Antimicrob Agents Chemother* 56, 3898–3904. doi: 10.1128/aac.06120-11

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). Fast tree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Reuter, S., Corander, J., de Been, M., Harris, S., Cheng, L., Hall, M., et al. (2015). Directional gene flow and ecological separation in Yersinia enterocolitica. *Microb Genom* 1:e000030. doi: 10.1099/mgen.0.000030

Riley, L. W. (2014). Pandemic lineages of extraintestinal pathogenic *Escherichia coli. Clin. Microbiol. Infect.* 20, 380–390. doi: 10.1111/1469-0691.12646

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

Seyfried, E. E., Newton, R. J., Rubert, K. F., Pedersen, J. A., and McMahon, K. D. (2010). Occurrence of tetracycline resistance genes in aquaculture facilities with varying use of oxytetracycline. *Microb. Ecol.* 59, 799–807. doi: 10.1007/s00248-009-9624-7

Stoesser, N., Sheppard, A. E., Pankhurst, L., De Maio, N., Moore, C. E., Sebra, R., et al. (2016). Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *mBio* 7:e02162. doi: 10.1128/mBio.02162-15

Sun, J., Chen, C., Cui, C. Y., Zhang, Y., Liu, X., Cui, Z. H., et al. (2019). Plasmid-encoded tet(X) genes that confer high-level tigecycline resistance in *Escherichia coli*. *Nat. Microbiol.* 4, 1457–1464. doi: 10.1038/s41564-019-0496-4

Tandogdu, Z., and Wagenlehner, F. M. (2016). Global epidemiology of urinary tract infections. *Curr. Opin. Infect. Dis.* 29, 73–79. doi: 10.1097/qco.0000000000000228

Tchesnokova, V. L., Rechkina, E., Larson, L., Ferrier, K., Weaver, J. L., Schroeder, D. W., et al. (2019). Rapid and extensive expansion in the United States of a new multidrug-resistant *Escherichia coli* clonal group, sequence type 1193. *Clin. Infect. Dis.* 68, 334–337. doi: 10.1093/cid/ciy525

Tchesnokova, V., Radey, M., Chattopadhyay, S., Larson, L., Weaver, J. L., Kisiela, D., et al. (2019). Pandemic fluoroquinolone resistant *Escherichia coli* clone ST1193 emerged *via* simultaneous homologous recombinations in 11 gene loci. *Proc. Natl. Acad. Sci. U.S.A.* 116, 14740–14748. doi: 10.1073/pnas.1903002116

Valenza, G., Werner, M., Eisenberger, D., Nickel, S., Lehner-Reindl, V., Holler, C., et al. (2019). First report of the new emerging global clone ST1193 among clinical isolates of extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* from Germany. *J. Glob. Antimicrob. Resist* 17, 305–308. doi: 10.1016/j.jgar.2019.01.014

Vila, J., Sáez-López, E., Johnson, J. R., Römling, U., Dobrindt, U., Cantón, R., et al. (2016). *Escherichia coli*: an old friend with new tidings. *FEMS Microbiol. Rev.* 40, 437–463. doi: 10.1093/femsre/fuw005

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

Wu, J., Lan, F., Lu, Y., He, Q., and Li, B. (2017). Molecular characteristics of ST1193 clone among phylogenetic group B2 non-ST131 fluoroquinolone-resistant *Escherichia coli*. *Front. Microbiol.* 8:2294. doi: 10.3389/fmicb.2017.02294

Wu, Y., Liu, C., Li, W. G., Xu, J. L., Zhang, W. Z., Dai, Y. F., et al. (2019). Independent microevolution mediated by mobile genetic elements of individual Clostridium difficile isolates from clade 4 revealed by whole-genome sequencing. *mSystems* 4:e252-18. doi: 10.1128/mSystems.00252-18

Xia, L., Liu, Y., Xia, S., Kudinha, T., Xiao, S. N., Zhong, N. S., et al. (2017). Prevalence of ST1193 clone and IncI1/ST16 plasmid in E. coli isolates carrying blaCTX-M-55 gene from urinary tract infections patients in China. *Sci. Rep.* 7:44866. doi: 10.1038/srep44866

Zhou, Z., Alikhan, N. F., Mohamed, K., Fan, Y., and Achtman, M. (2020). The enterobase user's guide, with case studies on *Salmonella* transmissions, Yersinia pestis phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* 30, 138–152. doi: 10.1101/gr.251678.119