



ORIGINAL ARTICLE

Molecular Epidemiological Surveillance of *bla*CTX-M in *Escherichia coli* from Indonesian Environmental Sources Based on Public Genomic Data

Ismail Shaleh Sikumbang¹, Marito Ade Lina Nasution², Muhammad Faqihuddin Siregar³

1,2,3 Independent Researcher, Padangsidempuan, North Sumatera, Indonesia

E-mail Corresponding: ismailskb08@gmail.com

ABSTRACT

Background: Antimicrobial resistance mediated by the *bla*CTX-M gene in *Escherichia coli* represents a major global public health threat, while molecular epidemiological surveillance data from environmental sources in Indonesia remain highly limited. This study aimed to analyze the molecular epidemiology of the *bla*CTX-M gene in environmental *Escherichia coli* isolates from Indonesia using a genomic surveillance approach based on publicly available genome data and Linux-based bioinformatics pipeline.

Method: This quantitative descriptive study employed a comparative genomics and phylogenomic analysis design based on an in silico approach. The study included 86 *Escherichia coli* genome assemblies retrieved from the Enterobase database using the selection criteria of Environment as the source niche and Indonesia as the country of origin. Data analysis was performed using QUAST, Prokka, AMRFinderPlus, ABRicate, MLST, Snippy, IQ-TREE, and FastTree on the Ubuntu 22.04 LTS operating system.

Result: All isolates met the assembly quality criteria, and the analytical pipeline was validated using both positive and negative controls. Most isolates originated from water and river samples, with the highest number collected in 2019. The *bla*CTX-M subtypes were predominantly assigned to the CTX-M-1 group (88,2%), particularly CTX-M-15 and CTX-M-55, while resistome profiling revealed a multidrug-resistant pattern. Sequence type ST2 and ST471 were the most prevalent, and although the genomic population was highly heterogeneous, four clonal clusters were identified.

Conclusion: These findings highlight the need for sustained deposition of environmental *Escherichia coli* genome data from Indonesia with more balanced temporal and geographic representation to strengthen the national One Health-based genomic surveillance system for antimicrobial resistance.

Keywords: *bla*CTX-M; *Escherichia coli*; environment; antimicrobial resistance; genomic surveillance



© 2026 e-SEHAD.

e-Sehad is an Open Access Journal. Published by Center Of Excellence Scientific Of Environmental And Health Diseases Universitas Jambi.

This is an open access article under the CC BY-NC-SA license

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

INTRODUCTION

Antimicrobial resistance (AMR) has become one of the most serious global health threats of the 21st century. The World Health Organization (WHO) has designated AMR as one of the top ten public health threats facing humanity. The latest data show that as many as 4.71 million deaths globally in 2021 were associated with antimicrobial-resistant bacterial infections, with 1.14 million deaths directly attributable to AMR, surpassing deaths from tuberculosis, malaria, and HIV/AIDS individually.(1,2) Current projections estimate that between 2025 and 2050 there will be 39 million deaths due to AMR, making it a global threat that requires an immediate and comprehensive response from all stakeholders in human, animal, and environmental health.(3)

The increasing prevalence of Extended Spectrum Beta-Lactamase (ESBL)-producing bacteria has become a major concern in global AMR epidemiology. ESBL is an enzyme that hydrolyzes the beta-lactam ring in penicillins and third-generation cephalosporins, causing resistance to nearly all classes of beta-lactam antibiotics except carbapenems.(4) The prevalence of ESBL-producing *Escherichia coli* varies between 20-70% across various regions of Asia and continues to increase along with the expanding use of third-generation antibiotics.(5) In Indonesia, ESBL-*E. coli* bacteria have been detected in various sources, ranging from clinical samples, livestock, to aquatic environments, with prevalence ranging from 9.9% to 86.1% depending on the sample type and collection location.(6)

Among the various ESBL types, the *bla*CTX-M gene dominates the global epidemiology of beta-lactam resistance. This gene is divided into several main groups, namely CTX-M1, CTX-M-2, CTX-M-8, CTX-M-9, and CTX-M-25, with the subtypes CTX-M-15 and CTX-M-14 being the most prevalent worldwide.(7) Genomic analysis of more than 2,144 *bla*CTX-M-55-carrying *E. coli* genomes shows endemic spread across nearly every continent, with the highest concentrations in China, the United States, and Thailand.(8) In *E. coli* isolates from broiler chickens in Indonesia, the *bla*CTX-M gene was detected in 97.5% of ESBL isolates, indicating the absolute dominance of this gene as the primary determinant of beta-lactam resistance in the local context.(9)

Escherichia coli has been internationally recognized as a primary indicator bacterium for antimicrobial resistance in the environment. As a commensal of the intestines of humans and warm-blooded animals that is ubiquitous, *E. coli* is easily found in various environmental compartments ranging from surface water, soil, to sediment, and its presence reflects the antibiotic selection pressure from human and animal activities in the surrounding area.(10) A genomic study of ESBL-producing *E. coli* in the Jakarta area found evidence of cross-sectoral One Health transmission between humans, broiler chickens, and surface water,(11) while in East Java, dairy cattle farm wastewater has also been proven to be a source of ESBL-producing *E. coli* with significant potential for environmental contamination.(12)

The environment serves as a reservoir for antibiotic resistance genes (ARGs), including

*bla*CTX-M. River water, domestic water, hospital wastewater, and sediment are environmental matrices rich in ARGs due to unmanaged inputs of human and animal feces as well as pharmaceutical industry waste.(13) Indonesian river waters affected by agricultural, livestock, and dense settlement activities have become a significant medium for the spread of ARGs, exacerbated by weak wastewater treatment infrastructure and sanitation in many regions.(10,14)

The spread of *bla*CTX-M in nature occurs mainly through horizontal gene transfer (HGT) mechanisms mediated by mobile genetic elements (MGEs) such as conjugative plasmids, transposons, and integrons.(15) IncF group plasmids are the most dominant vectors in the spread of *bla*CTX-M due to their high conjugation capability and their capacity to carry multiple ARGs simultaneously, thereby contributing to multidrug-resistance profiles.(16,13) A conjugation study in Tanzania also demonstrated that most plasmids carrying *bla*CTX-M-15 are IncF plasmids that are conjugatively transferred together with non-beta-lactam resistance genes.(17)

The potential for transmission of antibiotic resistance from the environment to humans through the One Health pathway has been well documented. A comprehensive genomic study in Jakarta using Whole Genome Sequencing (WGS) showed the possibility of transmission of ESBL-producing *E. coli* between humans, broiler chickens, and surface water, with ST410 isolates carrying *bla*CTX-M-15 distributed across sectors.(11) The

Indonesian government has established the National Action Plan for Antimicrobial Resistance Control (RAN-PRA), which integrates a cross-sectoral One Health approach spanning humans, animals, and the environment,(18) however, the implementation of environment-based AMR surveillance is still very limited, so molecular epidemiology data from environmental sources are not yet comprehensively available.(14)

The development of WGS technology has revolutionized the molecular epidemiological surveillance of pathogenic bacteria and AMR, enabling comprehensive resistome characterization, Sequence Type (ST) determination through Multi-Locus Sequence Typing (MLST), genetic relatedness analysis based on Single Nucleotide Polymorphism (SNP), and phylogenetic reconstruction in one integrated analysis.(19) WGS has now become the backbone of global One Health surveillance, as implemented by WHO through the GLASS system, which covers 130 countries.(14) The use of public genomic databases, particularly EnteroBase, which integrates more than one million bacterial genomes, opens up great opportunities in large-scale genomic data-based AMR surveillance without requiring de novo sequencing.(20,21) An EnteroBase-based study from broiler chicken farms in the United Kingdom successfully tracked the spread of ESBL-producing *E. coli* from farms to humans using a cgMLST-based phylogenetic tree, proving the high utility of this platform for cross-sectoral genomic surveillance.(22)

Although genomic data on *E. coli* from Indonesian have been stored in public databases

such as NCBI and EnteroBase, to date there has been no research that specifically extracts, analyzes, and integrates the entire genomic dataset of Indonesian environmental *E. coli* carrying *bla*CTX-M using a Linux-based bioinformatics approach. As a result, the resistome profile, *bla*CTX-M subtype distribution, MLST patterns, phylogenetic relationships, and genomic cluster dynamics of Indonesian environmental isolates have not been comprehensively mapped.(23,19) Given the increasing threat of AMR, this study aims to analyze the molecular epidemiology of the *bla*CTX-M gene in *Escherichia coli* originating from the Indonesian environment through a genomic surveillance approach based on public genomic data analyzed using a comprehensive bioinformatics pipeline, in order to support national One Health policy in AMR control.

METHOD

Type of Research

This research is a descriptive quantitative study that systematically and measurably describes the characteristics of the research object based on numerical data, as it analyzes secondary data in the form of genomic sequences available in public databases in an objective and structured manner.(24) This study adopts a molecular epidemiology approach, which is a branch of science that integrates molecular biology with epidemiology to understand the distribution, etiology, and risk factors of disease at the molecular level,(25) thereby enabling the tracing of the origin and spread of pathogens

with greater precision compared to conventional epidemiological methods.

An *in silico* bioinformatics study is used as the main strategy, namely a computational-based method that utilizes software and biological databases to analyze genomic information without wet laboratory testing, making it more efficient in terms of time and cost.(26,27) The research design specifically uses a comparative genomics study approach that compares the sequences and genomic structures of various isolates to identify genetic similarities and differences that are evolutionarily meaningful,(28) complemented by phylogenomic analysis, namely the reconstruction of phylogenetic trees using genome-scale data to trace evolutionary relationships more accurately compared to single-gene analysis, as genomics-based phylogenetic methods have been shown to produce tree topologies with better resolution and nodal support.(29)

Data Source and Sample Selection

Data collection was conducted online through the EnteroBase platform (<https://enterobase.warwick.ac.uk>), while data analysis was carried out using the Linux Ubuntu 22.04 LTS command-line-based operating system on local hardware, through four sequential stages of activity over four months, namely proposal preparation, data collection, data analysis, and final report writing.

The study population consisted of all *Escherichia coli* genome sequences stored in the EnteroBase webserver database. The study sample consisted of *Escherichia coli* genome

assemblies filtered based on two metadata attributes, namely source niche "Environment" and country "Indonesia," aiming to obtain isolates that are geographically and ecologically relevant to the research context. Inclusion criteria included the availability of downloadable genome assemblies in FASTA format, completeness of country-of-origin and source niche metadata, and confirmation of Indonesian environmental origin. Exclusion criteria included corrupted or unprocessable genome assemblies, duplicate genomes, incomplete or ambiguous metadata, and genomes indicating high contamination based on assembly quality parameters.

Bioinformatics Instruments and Tools

This study used a laptop with specifications of 2 GB RAM, a 2-core CPU, and 50 GB of storage running Ubuntu 22.04 LTS, consistent with reports that medium-scale bacterial genomic analysis can be run on desktop computers with limited resources using conda/mamba-based pipelines.(30) The bioinformatics tools used included QUAST and BUSCO for assembly quality evaluation, Prokka for automated prokaryotic genome annotation,(31) AMRFinderPlus and ABRicate for the detection of antimicrobial resistance genes,(32) the mlst tool for sequence type (ST) determination, as well as Snippy and Snippy-core for variant calling and the construction of core-genome alignments that served as input for IQ-TREE and FastTree in building maximum-likelihood phylogenetic trees, with the results visualized using FigTree.

The positive control used was the reference genome of *Escherichia coli* JJ1886 (CP006784.1), namely an *E. coli* ST131 subclone H30-Rx carrying *bla*CTX-M-15 that has been scientifically confirmed,(33) with validation thresholds of identity $\geq 90\%$ and coverage $\geq 60\%$.(34) The negative control used the reference genome of *Escherichia coli* strain NCTC86 (CP019778), namely an original isolate from the pre-antibiotic era that does not carry any beta-lactam resistance genes or ESBL,(35) as is commonly used in WGS-based resistance genomic studies to ensure pipeline specificity and prevent false positives.(36)

Bioinformatics Analysis Procedure

Genome assembly quality was evaluated using QUAST v5.3.0 and BUSCO v6.1.0, with parameters of contig number ≤ 500 , N50 value $\geq 50,000$ bp, total genome length of 4.5-5.8 Mb, GC content of approximately 50-51%, and BUSCO completeness $\geq 95\%$,(37,38,39) referring to the standard that a higher N50 value reflects better assembly contiguity.(40) Genomes that did not one more parameters were eliminated from the analysis pipeline.(41)

Genome annotation was performed using Prokka v1.15.6 with parameters specific to *Escherichia coli*.(31) Resistome analysis was performed using AMRFinderPlus v4.2.7 and ABRicate v1.4.0 with the NCBI AMR, CARD, and ResFinder databases, using identity thresholds $\geq 90\%$ and coverage $\geq 50\%$ for AMRFinderPlus, and identity $\geq 80\%$ and coverage $\geq 80\%$ for ABRicate.(42,43) Identification of CTX-M subtypes was performed using ABRicate with the ResFinder and CARD databases, with

coverage thresholds $\geq 60\%$ and identity $\geq 90\%$ as used in WGS-based *bla*CTX-M identification studies in *E. coli*.(44)

Multi-Locus Sequence Typing (MLST) analysis was performed using the *mlst* v2.33.1 tool with the Achtman scheme (PubMLST) targeting eight housekeeping genes (*dinB*, *icdA*, *pabB*, *putP*, *trpA*, *trpB*, and *uidA*) to determine ST.(45,46) SNP and core genome analysis was performed using *Snippy* v3.1 against the *Escherichia coli* K-12 MG1655 reference, followed by *Snippy-core* to generate a core SNP alignment from all isolates.(47) Phylogenomic reconstruction was performed using *IQ-TREE* with the GTR-G model and 1,000 ultrafast bootstrap replicates, along with *FastTree* for topology validation, a bootstrap value ≥ 70 was used as the acceptable clade threshold, and ≥ 95 indicated a statistically strong clade.(48,49,50)

Isolates with a genomic SNP distance ≤ 20 were defined as clonal and considered potential transmission clusters.(8) Data visualization was performed using *FigTree* v1.4.4 for phylogenetic trees, *R* v4.3.2 with the *ggplot* and *pheatmap* packages for the resistome and SNP matrix with hierarchical clustering using the Ward method,(51) and *Python* v3.9.23 with the *pandas*, *seaborn*, and *matplotlib* libraries for visualizing ST distribution and genomic metadata integration.(23) Pipeline validity was tested through the sensitivity of *bla*CTX-M detection on a random subset of 20 genomes, the specificity of *Prokka* annotation against the *Escherichia coli* K-12 MG1655 reference, and the concordance rate of *RGI/CARD* and *ABRicate*

detection, with an expected agreement of $\geq 9\%$.(52)

Data Analysis

Descriptive analysis included the calculation of the frequency distribution and proportion of all *bla*CTX-M subtypes, the distribution of STs based on the eight-gene PubMLST MLST scheme,(45) as well as the distribution of the resistome representing the diversity of ARG classes in each isolate, presented in graphs using *R* and *Python*. Comparative genomic analysis was conducted to compare resistome profiles between isolates based on environmental source, ST distribution between groups, and the SNP distance matrix between pairs of isolates using *snp-dists*.

Phylogenomic analysis included core SNP analysis to identify single nucleotide variants in core regions, clade analysis to map monophyletic groupings based on the maximum-likelihood (ML) tree, and genomic relatedness analysis to determine evolutionary distances between isolates and identify high-risk clusters, following the approach that has been applied in global-scale *E. coli* genomic surveillance.(53)

RESULT

Selection and Quality of Genome Assembly

Sample selection began with the screening of *Escherichia coli* genomes in the Enterobase database based on the criteria of source niche "Environment" and country "Indonesia," yielding 86 genome assemblies originating from Indonesian environmental sources. All selected genomes underwent

quality assessment using QUAST with parameters including contig number, N50

value, total genome length, and GC content, as presented in Figure 1.

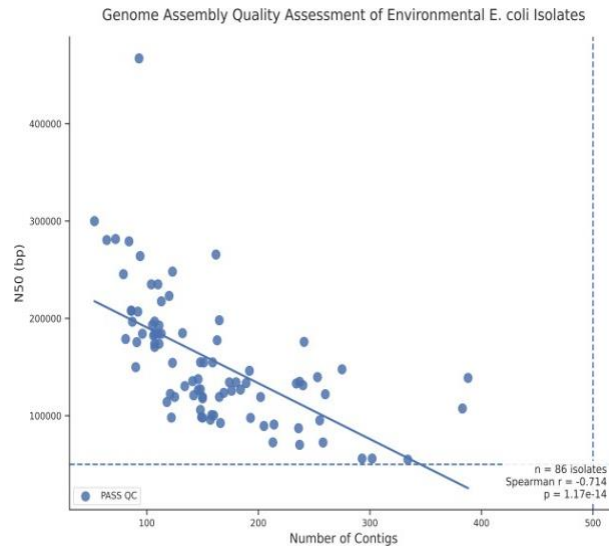


Figure 1. Distribution of N50 Values and Contig Numbers Across All Indonesian Environmental *Escherichia coli* Isolates Based on QUAST Evaluation

Figure 1 shows the results of assembly quality evaluation for 86 Indonesian environmental *E. coli* isolates based on contig number and N50 value. All isolates fell within a contig number range below 500, with most concentrated in the range of 50 to 250 contigs and N50 values ranging from 50,000 to 300,000 bp, such that all isolates met the established upper limit for contig number and minimum N50 threshold. The regression trend line showed a negative relationship between contig number and N50 value, confirmed through a Spearman correlation test with a coefficient r of -0.714 and a p -value of 1.17×10^{-14} , indicating a statistically significant relationship. This pattern indicates that

isolates with fewer contig tend to have higher N50 values, reflecting better assembly contiguity. Overall, all isolates were declared to have passed quality control (PASS QC) and could be used in the subsequent bioinformatics analysis stage.

Validation of Positive and Negative Controls

Validation of the bioinformatics pipeline was performed using two reference genomes, namely *Escherichia coli* JJ1886 as the positive control carrying the *bla*CTX-M-15 gene, and *Escherichia coli* strain NCTC86 as the negative control not carrying *bla*CTX-M or any other ESBL genes. The validation results are presented in Table 1.

Table 1. Validation Results of the Bioinformatics Pipeline Using *Escherichia coli* JJ1886 (Positive Control) and *Escherichia coli* strain NCTC86 (Negative Control)

Control	Strain	<i>bla</i> CTX-M Detected	Other ESBL Detected	Validation Result
Positive	JJ1886	<i>bla</i> CTX-M-15	-	Passed
Negative	NCTC86	-	-	Passed

Based on Table 1, the positive control *Escherichia coli* JJ1886 showed detection of the *bla*CTX-M-15 gene and no other ESBL genes were found, so it was declared to have passed validation because the pipeline successfully identified the ESBL gene that should have been detected. Conversely, the negative control *Escherichia coli* strain NCTC86 showed no detection of *bla*CTX-M or any other ESBL genes, so it was also declared to have passed validation because there was no false positive detection. The successful validation of both controls simultaneously indicates that the

bioinformatics pipeline used is valid and suitable for use in resistome analysis of all *E. coli* genome samples originating from the Indonesian environment.

Temporal and Geographic Distribution of Indonesian Environmental *Escherichia coli* Genomes

The genome search results show that *E. coli* isolates originating from the Indonesian environment came from various sources with an uneven distribution, as presented in Figure 2.

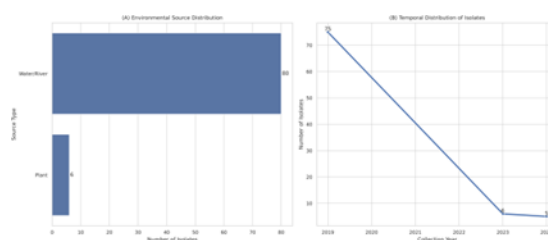


Figure 2. Distribution of Environmental Sources (a) and Temporal Distribution (b) of Indonesian Environmental *Escherichia coli* Isolates from the EnteroBase Database

Based on Figure 2 part (a), the distribution of isolate environmental sources was dominated by isolate originating from water and river sources, totalling 80 isolates, while isolates originating from plant sources numbered only 6, indicating that bodies of water, particularly rivers, are the dominant environmental reservoir for the presence of *E. coli* in the public genomic data analyzed. In Figure 2 part (b), the temporal distribution of isolates shows a pattern of decreasing isolate numbers from year to year. The highest number of isolates was recorded in 2019 (75 isolates), then sharply declined to 6 isolates in 2023 and

5 isolates in 2024, illustrating that the availability of Indonesian environmental *E. coli* genome data in public databases is still highly dependent on specific collection periods.

Resistome Profile of Indonesian Environmental *Escherichia coli*

Resistome analysis was performed on all isolates using AMRFinderPlus and ABRicate, and the detection results were visualized in the form of a binary heatmap showing the presence or absence of resistance genes in each isolate, as presented in Figure 3.

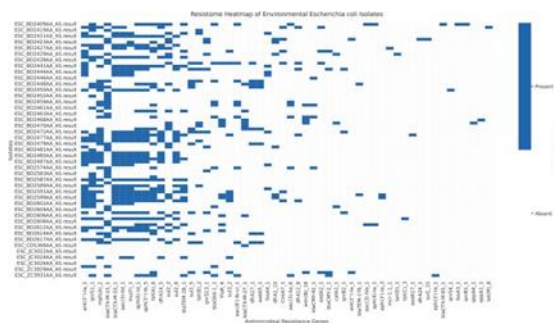


Figure 3. Heatmap of the Resistome Profile of Indonesian Environmental *Escherichia coli* Isolates – Distribution of Antibiotic Resistance Genes (ARGs) per Isolate Based on AMRFinderPlus and ABRicate Analysis

Figure 3 shows that most isolates carried more than one antibiotic resistance gene, with a diverse distribution pattern among isolates. A number of genes appeared dominant and were distributed across nearly all isolates, including beta-lactam genes (*bla*CTX-M-15, *bla*CTX-M-55, and *bla*TEM-1B), aminoglycoside genes (*aac*(3)IId, *aph*(6)-Id, and *aph*(3'')-Ib), as well as tetracycline (*tet*(A)) and sulfonamide (*sul*1, *sul*2) genes, indicating a high selection pressure from these antibiotic classes in the isolates environment of origin. In addition, resistance genes were also found in certain subsets of isolates, such as fluoroquinolone genes (*qnr*S1, *qnr*B2, *aac*(6')-Ib-cr), phenicol genes (*cat*A1, *cml*A7), macrolide genes (*erm*(B), *mph*(A)), and the *mcr*-1.1 gene associated with

colistin resistance, indicating that resistance to last-line antibiotics has been found, albeit at a relatively low frequency. The heatmap pattern generally shows the clustering of isolates based on the similarity of their resistance gene profiles, with some isolates exhibiting very broad resistance profiles while other isolates carried few or no detectable ARGs.

Identification and Distribution of *bla*CTX-M Subtypes

Identification of *bla*CTX-M subtypes was carried out to determine the genetic variation across all isolates. The analysis results show that all isolates carried the *bla*CTX-M gene, which was grouped into two main groups, namely CTX-M-1 and CTX-M-9, as presented in Figure 4.

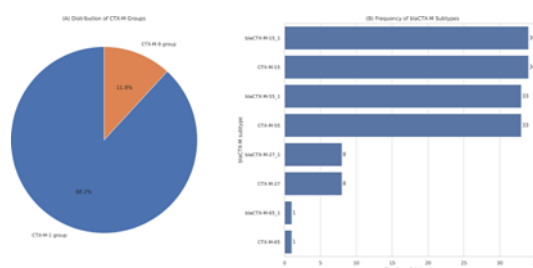


Figure 4. Distribution of *bla*CTX-M Subtypes in Indonesian Environmental *Escherichia coli* Isolates – (a) Proportion of CTX-M Groups, (b) Frequency of Specific Subtypes

The identification results show that the CTX-M-1 group dominated the distribution of *bla*CTX-M in Indonesian environmental isolates

with a proportion of 88.2%, while the CTX-M-9 group was found in only 11.8% of isolates (Figure 4a). At the subtype level, the CTX-M-1

group was represented by three subtypes, namely CTX-M-15, CTX-M-55, and CTX-M-27, with detection counts of 34, 33, and 8 times, respectively, across different isolates (Figure 4b), while the CTX-M-9 group was represented by only one subtype, namely CTX-M-65, which was detected in one isolate. Some isolates showed the *bla*CTX-M gene in two copies or slightly different alleles within the same genome, so they were recorded as separate entries by the resistance gene annotation database. Overall, the dominance of the CTX-M-15 and CTX-M-55 subtypes indicates that these

two variants are the most widely circulating *bla*CTX-M subtypes in Indonesian environmental *E. coli* compared to other subtypes.

Distribution of Sequence Type (ST) Based on MLST Analysis

MLST analysis based on the Achtman scheme was performed to identify STs across all isolates. Figure 5 displays the distribution of the twenty STs with the highest frequencies found.

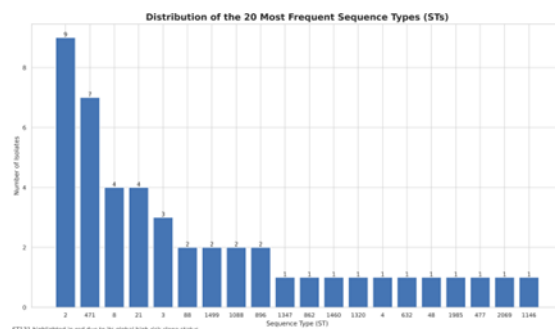


Figure 5. Distribution of Sequence Type (ST) in Indonesian Environmental *Escherichia coli* Isolates Based on the Achtman MLST Scheme – 20 STs with the Highest Frequencies

Bases Figure 5, ST2 was the most dominant sequence type with nine isolates, followed by ST471 with seven isolates. These two STs accounted to other STs, indicating the dominance of certain clones within the isolate population studied. Furthermore, ST8 and ST21 were each found in four isolates, ST3 in three isolates, and ST88, ST1499, ST1088, and ST896 each in two isolates. Eleven other STs, ST1347, ST862, ST1460, ST1320, ST4, ST632, ST48, ST477, and ST1146, were each found in only one isolate. This distribution pattern illustrates a combination of widespread dominant clones and various sporadic single STs, reflecting the

genetic diversity of the Indonesian environmental *E. coli* population.

SNP and Core Genome Analysis

SNP and core genome analysis was performed using Snippy v3.1 against the *Escherichia coli* K-12 MG1655 reference genome, followed by Snippy-core to generate a core SNP alignment from all 86 isolates. The pairwise SNP distance matrix between isolates was calculated using snp-dists and visualized as a heatmap with hierarchical clustering in Figure 6, while the descriptive statistics are presented in Table 2.

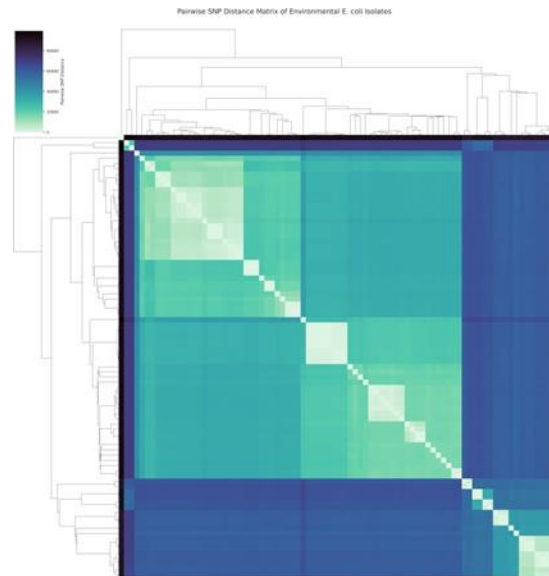


Figure 6. Heatmap and Hierarchical Clustering Based on the Pairwise SNP Distance Matrix Between Indonesian Environmental *Escherichia coli* Isolates

The heatmap in Figure 6 shows the distribution of pairwise SNP distances among the 86 isolates through a color gradient, along with hierarchical clustering dendrograms on both sides. Most of the heatmap area displays light to very light colors, reflecting high SNP distances among most pairs or isolates, indicating that the population is genomically highly heterogeneous without the dominance

of a single clone. Nevertheless, there are several small areas with darker colors and localized distribution, suggesting the presence of pairs or small groups of isolates with low SNP distances that potentially indicate a clonal relationship. This pattern is consistent with the high median SNP distance value (33,645.5) as reflected in Table 2.

Table 2. Core SNP Statistics and Identification of Clonal Clusters (SNP Distance ≤ 20) in Indonesian Environmental *Escherichia coli* Isolates

Parameter	Value	Description
Number of isolates	86	Total isolates analyzed
Minimum SNP distance	2	Closest pairs of isolates
Median SNP distance	33,645.5	Median of all pairs isolates
Mean SNP distance	41,392.99	Average of all pairs of isolates
Maximum SNP distance	99,504	Farthest pair of isolates
Number of SNP pairs ≤ 20	13	Candidate clonal relationship
Number of clonal clusters	4	Clusters with SNP distance ≤ 20
Number of isolates in the largest cluster	4	Number of isolates in the largest cluster

Table 2 presents the descriptive statistics of the core SNP from the 86 isolates. The minimum SNP distance value of 2 indicates that there is at least one pair of isolates that is genomically almost identical, while the

maximum value reaches 99,504, indicating the presence of a pair of isolates with very large genomic differences. The wide and asymmetric median (33,645.5) and mean (41,392.99) values (mean higher than median/right

skewed) reinforce the picture that the isolate population is not dominated by a single clone. Based on the SNP distance threshold of ≤ 20 as the clonality criterion, 13 pairs of isolates were identified that met this criterion, forming 4 clonal clusters with the largest cluster consisting of 4 isolates, indicating the presence of clear localized clonal transmission events among some of the isolates.

Phylogenomic and Genomic Cluster Analysis

Phylogenomic analysis was performed on all isolates that met genome quality criteria using a Maximum-Likelihood approach based on the core genome SNP matrix, with each isolate annotated according to its ST from the Achtman MLST scheme and its *bla*CTX-M subtype from the CARD database-based resistome analysis, as presented in Figure 7.

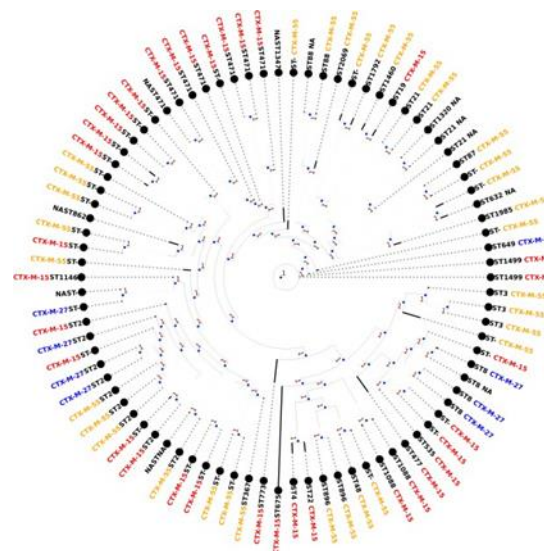


Figure 7. Maximum-Likelihood Phylogenomic Tree of Indonesian Environmental *Escherichia coli* Isolates – Annotated with Sequence Type (ST) and *bla*CTX-M Subtype

The phylogenomic tree in Figure 7 is visualized in circular format, with color annotations of red for CTX-M-15, orange for CTX-M-55, blue for CTX-M-27, and black for isolates without subtype annotation (NA). The tree shows a verified topology with several main clusters that can be clearly distinguished. High ST diversity is evident in the distribution of annotations across all branches of the tree, including ST2, ST3, ST8, ST21, ST88, ST410, ST471, ST535, ST649, ST773, ST775, ST862, ST896, ST1086, ST1088, ST1146, ST1320, ST1460, ST1499, ST1792, ST1985, ST2069, and several isolates with undefined ST (NA),

indicating that the population is polyphyletic and not dominated by a single clone.

The color annotation shows that the CTX-M-55 subtype has the widest distribution, found in nearly all of the tree's main clusters without forming a single exclusive monophyletic cluster. The CTX-M-15 subtype also shows a wide distribution, including ST410 and ST1466 in the upper left cluster as well as in ST1499, ST535, ST1086, and ST1088 in the lower right cluster, while the CTX-M-27 subtype was found at a lower frequency, mainly concentrated in ST2 isolates and several ST8 isolates. The distribution pattern of these three

subtypes, which does not follow a single phylogenetic cluster, indicates repeated and independent acquisition of the *bla*CTX-M gene through horizontal gene transfer across various different ST genetic backgrounds. A cluster composed of ST2 isolates with CTX-M-15, CTX-M-27, and CTX-M-55 annotations side by side shows subtype diversity within the same clonal background, with high bootstrap values (approaching 1) at several internal branches confirming the statistical reliability of this cluster grouping.

DISCUSSION

The successful selection and validation of 86 *E. coli* isolates from Indonesian environmental sources in the EnteroBase database, with all isolates passing QUASt quality control, demonstrates that the developed bioinformatics pipeline is capable of working consistently and reproducibly. The dominance of isolates from water and river sources (80 isolates/93.0%) compared to plant sources (6 isolates/7.0%) does not merely reflect the ecological preference of *E. coli*, but also indicates that bodies of water are the primary convergence point of ARG input from various anthropogenic sources, including domestic waste, agriculture, and livestock farming. This is consistent with the findings of Maric et al., who reported ESBL-*E. coli* detected in 61% of water samples compared to only 35% of sediment samples, confirming that open waters are the most genetically diverse active reservoir.(54) Meanwhile, the highly uneven temporal distribution, with 75 isolates (87.2%) originating from 2019 and only 11 isolates from

2023-2024, indicates a low deposition rate of Indonesian environmental genomic data into public databases following the COVID-19 pandemic, which indirectly limits the ability of research to capture the current dynamics of *bla*CTX-M spread in Indonesian aquatic environments.

The resulting resistome profile shows that Indonesian environmental *E. coli* isolates not only carry the *bla*CTX-M gene, but also consistently carry resistance genes to aminoglycosides, tetracyclines, sulfonamides, and *bla*TEM-1B. This multiclass co-resistance pattern is not a random occurrence, but rather a direct consequence of the role of IncF group conjugative plasmids as vectors that simultaneously carry multiple ARGs within a single replicon unit. From a public health perspective, these findings mean that infections caused by these isolates are not only resistant to third-generation cephalosporins, but also to most first-line antibiotics available in Indonesian primary healthcare facilities. Even more concerning is the discovery of the *mcr*-1.1 gene in a small number of isolates, indicating that resistance to colistin, a last-line antibiotic, has emerged in Indonesian environmental compartments, well before it might potentially be detected clinically. Richter et al. reported similar findings in which 95.1% of environmental ESBL-*E. coli* isolates from South Africa carried resistance to three or more antibiotic classes, confirming that the complexity of multidrug-resistance in environmental *E. coli* is a global phenomenon no longer limited to the clinical context.(55)

The finding that all isolates carried only two *bla*CTX-M groups, namely CTX-M-1 (88.2%) and CTX-M-9 (11.8%), with the subtypes CTX-M-15 and CTX-M-55 as the most dominant, has important epidemiological implications. The dominance of the CTX-M-1 group in Indonesian aquatic environments is consistent with the global pattern in which this group, particularly CTX-M-15, is associated with broad-host-range IncF plasmids with high conjugative transfer capacity. Yang et al. confirmed that *bla*CTX-M-55 has spread endemically across nearly every continent through a combination of clonal transmission and horizontal plasmid transfer, with the highest concentration in the Asian region.(8) The fact that all three subtypes (CTX-M-15, CTX-M-55, and CTX-M-27) were found scattered across various different sequence types on the phylogenomic tree, rather than concentrated within a single monophyletic cluster, is strong evidence that the spread of *bla*CTX-M in the Indonesian environment is dominated by horizontal gene transfer mechanisms, rather than single clonal expansion. This pattern has serious consequences for control strategies, as clone-control-based interventions alone will not be sufficient without also cutting off plasmid transfer pathways in the environment.

MLST analysis identified ST2 (9 isolates) and ST471 (7 isolates) as the most dominant sequence types, differing from reports of global high-risk clones such as ST131 and ST410, which generally dominate clinical isolates and densely populated urban environments. Puspadari et al. who studied

ESBL-*E. coli* in Jakarta, found ST410 to be the dominant ST in environmental isolates, while ST131 dominated human isolates.(11) The difference from the findings of this study likely reflects differences in ecological selection pressure between more open general river waters and densely populated urban areas intensively exposed to clinical and domestic waste. Maric et al. also confirmed that ST diversity in environmental waters is far higher than in clinical isolates, with 134 different STs found among 564 water and sediment isolates, underscoring that surveillance based on clinical isolates alone is unable to capture the full genetic diversity of *bla*CTX-M-carrying *E. coli* circulating in Indonesian aquatic ecosystems.(54)

The very high median SNP distance value (33,646.5) in the pairwise matrix of all isolates confirms that the Indonesian environmental *E. coli* population is highly heterogeneous genomically, without a single dominant clone spreading widely. However, the identification of 4 clonal clusters from 13 pairs of isolates with an SNP distance ≤ 20 is an epidemiologically meaningful finding, indicating the presence of clear localized transmission events, likely originating from the same contamination source or geographically adjacent sample collection sites. Fagerstrom et al. showed that ST131 isolates from river water can be genomically nearly identical to clinical isolates from urinary tract infection patients, with differences of ≤ 10 cgMLST alleles, proving that clonal clusters in environmental isolates are an epidemiological signal that should not be ignored.(56) Overall, the integration of

resistome, MLST, SNP, and phylogenomic results in this study provides a comprehensive molecular epidemiological picture of *bla*CTX-M in Indonesian environmental *E. coli*, while also affirming the urgency of environmental genomic monitoring as an integral component of the national One Health-based AMR surveillance system, as mandated in Indonesia's National Action Plan for Antimicrobial Resistance Control.

CONCLUSION

This study shows that the distribution of *bla*CTX-M-carrying *E. coli* genomes from the Indonesian environment is uneven, both in terms of source and collection period, with most isolates originating from water and rivers and concentrated in 2019. Variation in *bla*CTX-M subtypes consisted of the CTX-M-1 (88.2%) and CTX-M-9 (11.8%) groups, with CTX-M-15 and CTX-M-55 as the most dominant subtypes, while the resistome profile showed that most isolates were multidrug-resistant, with dominant genes belonging to the beta-lactam, aminoglycoside, tetracycline, and sulfonamide classes, as well as the discovery of the last-line resistance gene *mcr*-1.1, albeit at a low frequency. The sequence type distribution showed high diversity, with ST2 and ST471 being the most dominant, while the genetic relatedness between isolates was

heterogeneous, with a median SNP distance of 33,646.5 without the dominance of a single clone, although 4 clonal clusters (SNP ≤ 20) were identified, indicating localized clonal transmission. The genomic cluster pattern showed that the CTX-M-55 and CTX-M-15 subtypes were widely distributed across various phylogenetic clusters, while CTX-M-27 was more concentrated in ST2 and ST8, indicating repeated *bla*CTX-M acquisition through horizontal gene transfer across different ST backgrounds.

Based on these findings, it is recommended to increase the continuous deposition of Indonesian environmental *E. coli* genomic data into public databases with a more even temporal and geographic distribution, accompanied by phenotypic verification through antibiotic susceptibility testing and experimental conjugation assays, periodic updates to resistance gene annotation databases, completeness of location and field epidemiology metadata, ongoing surveillance of last-line resistance genes such as *mcr*-1.1, and further studies on mobile genetic elements such as plasmids and transposons that play a role in the spread of the *bla*CTX-M gene between isolates, in order to support the strengthening of Indonesia's One Health-based National Action Plan for Antimicrobial Resistance Control.

REFERENCE

1. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet*. 2024;404(10459):1199-1226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1)

2. Walsh TR, Gales AC, Laxminarayan R, Dodd PC. Antimicrobial resistance: addressing a global threat to humanity. *PLoS Med.* 2023;20(7):e1004264. <https://doi.org/10.1371/journal.pmed.1004264>
3. IHME. *Antimicrobial resistance (AMR)* [Internet]. Seattle (WA): Institute for Health Metrics and Evaluation; 2025 [cited 2026 May 30]. Available from: <https://www.healthdata.org/research-analysis/health-topics/antimicrobial-resistance-amr>
4. Sintondji K, Fabiyi K, Hougbenou J, Koudokpon H, Lègba B, Amoussou H, et al. Prevalence and characterization of ESBL-producing *Escherichia coli* in healthy pregnant women and hospital environments in Benin: an approach based on Tricycle. *Front Public Health.* 2023;11:1227000. <https://doi.org/10.3389/fpubh.2023.1227000>
5. Di Marcantonio L, Chiatamone Ranieri S, Toro M, Marchegiano A, Cito F, Sulli N, et al. Comprehensive regional study of ESBL *Escherichia coli*: genomic insights into antimicrobial resistance and inter-source dissemination of ESBL genes. *Front Microbiol.* 2025;16:1595652. <https://doi.org/10.3389/fmicb.2025.1595652>
6. Faridah HD, Wibisono FM, Wibisono FJ, Nisa N, Fatimah F, Effendi MH, et al. Prevalence of the blaCTX-M and blaTEM genes among extended-spectrum beta-lactamase-producing *Escherichia coli* isolated from broiler chickens in Indonesia. *J Vet Res.* 2023;67(2):179. <https://doi.org/10.2478/jvetres-2023-0025>
7. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β -lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother.* 2017;72(8):2145-2155. <https://doi.org/10.1093/jac/dkx146>
8. Yang JT, Zhang LJ, Lu Y, Zhang RM, Jiang HX. Genomic insights into global blaCTX-M-55-positive *Escherichia coli* epidemiology and transmission characteristics. *Microbiol Spectr.* 2023;11(4):e01089-23. <https://doi.org/10.1128/spectrum.01089-23>
9. Hidayatullah N, Suandy I, Intanon M, Alter T, Susanti O, Herpianti A, et al. Occurrence and molecular characterization of extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* in broilers in Indonesia. *Antibiotics (Basel).* 2025;14(10):1030. <https://doi.org/10.3390/antibiotics14101030>
10. Megantara I, Sylviana N, Amira PA, Pradini GW, Krissanti I, Lesmana R. Potential of waterbodies as a reservoir of *Escherichia coli* pathogens and the spread of antibiotic resistance in the Indonesian aquatic environment. *J Water Sanit Hyg Dev.* 2023;13(10):776-792. <https://doi.org/10.2166/washdev.2023.040>
11. Puspandari N, Pratama R, Goodman R, Suandy I, Efadeswarni E, Febriana T, et al. Genomic investigation of ESBL-producing *Escherichia coli* reveals likely transmission within and between One Health sectors in Jakarta, Indonesia. *Research Square* [Preprint]. 2025. <https://doi.org/10.21203/rs.3.rs-7825238/v1>
12. Dameanti FNAEP, Yanestria SM, Widodo A, Effendi MH, Plumeriastuti H, Tyasningsih W, et al. Incidence of *Escherichia coli* producing extended-spectrum beta-lactamase in wastewater of dairy farms in East Java, Indonesia. *Biodiversitas.* 2023;24(2):1143-1150. https://repository.unair.ac.id/127124/1/Artikel_15_Hani_Plumeriastuti.pdf
13. Kumavath R, Gupta P, Tatta ER, Mohan MS, Salim SA, Busi S. Unraveling the role of mobile genetic elements in antibiotic resistance transmission and defense strategies in bacteria. *Front Syst Biol.* 2025;5:1557413. <https://doi.org/10.3389/fsysb.2025.1557413>
14. World Health Organization. *Global Antibiotic Resistance Surveillance Report 2025: WHO Global Antimicrobial Resistance and Use Surveillance System (GLASS)* [Internet]. Geneva: World Health Organization; 2025. Available from: <https://www.who.int/publications/i/item/9789240116337>

15. Zhou W, Zhang E, Zhou J, He Z, Zhou Y, Han J, et al. Characterization and comparative genomics analysis of IncFII multi-resistance plasmids carrying blaCTX-M and type 1 integrons from *Escherichia coli*. *Front Microbiol.* 2021;12:753979. <https://doi.org/10.3389/fmicb.2021.753979>
16. Li P, Yan L, Song J, Lin C, Zeng F, Zeng S. Involvement of the blaCTX-M-3 gene in emergence of a peculiar resistance phenotype in *Klebsiella pneumoniae*. *Front Cell Infect Microbiol.* 2025;15:1545157. <https://doi.org/10.3389/fcimb.2025.1545157>
17. Minja CA, Shirima G, Mshana SE. Conjugative plasmids disseminating ctx-m-15 among humans, animals and the environment in Mwanza, Tanzania: a need to intensify One Health approach. *Antibiotics (Basel)*. 2021;10(7):836. <https://doi.org/10.3390/antibiotics10070836>
18. Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan Republik Indonesia. *Rencana aksi nasional pengendalian resistensi antimikroba* [Internet]. Jakarta: Kemenko PMK; 2022. Available from: https://cdn.who.int/media/docs/default-source/searo/indonesia/20221115_nap-on-amr.pdf
19. Vanstokstraeten R, Pierard D, Crombé F, De Geyter D, Wybo I, Muyldermans A, et al. Genotypic resistance determined by whole genome sequencing versus phenotypic resistance in 234 *Escherichia coli* isolates. *Sci Rep.* 2023;13(1):449. <https://doi.org/10.1038/s41598-023-27723-z>
20. Horesh G, Blackwell GA, Tonkin-Hill G, Corander J, Heinz E, Thomson NR. A comprehensive and high-quality collection of *Escherichia coli* genomes and their genes. *Microb Genom.* 2021;7(2):000499. <https://doi.org/10.1099/mgen.0.000499>
21. Murphy R, Palm M, Mustonen V, Warringer J, Farewell A, Parts L, et al. Genomic epidemiology and evolution of *Escherichia coli* in wild animals in Mexico. *mSphere.* 2021;6(1):e00738-20. <https://doi.org/10.1128/msphere.00738-20>
22. Duggett N, AbuOun M, Stubberfield E, Turner O, Randall L, Horton R, et al. Genomic surveillance of extended-spectrum cephalosporin-resistant *Escherichia coli* isolated from poultry in the UK from 2016 to 2020. *Front Microbiol.* 2024;14:1335173. <https://doi.org/10.3389/fmicb.2023.1335173>
23. Yehouenou CL, Bogaerts B, De Keersmaecker SCJ, Roosens NH, Marchal K, Tchiakpe E, et al. Whole-genome sequencing-based antimicrobial resistance characterization and phylogenomic investigation of 19 multidrug-resistant and extended-spectrum beta-lactamase-positive *Escherichia coli* strains collected from hospital patients in Benin in 2019. *Front Microbiol.* 2021;12:752883. <https://doi.org/10.3389/fmicb.2021.752883>
24. Sugiyono. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta; 2013. Available from: <https://www.scribd.com/document/671612229/Sugiyono-2013-Metode-Penelitian-Kuantitatif-Kualitatif-dan-R-D-1>
25. Bilal MY, Klutts JS. Molecular epidemiological investigations of localized SARS-CoV-2 outbreaks: utility of public algorithms. *Epidemiologia (Basel)*. 2022;3(3):402-411. <https://doi.org/10.3390/epidemiologia3030031>
26. Alizadehmohajer N, Zahedifar S, Sohrabi E, Shaddel Basir S, Nourigheimasi S, Falak R, et al. Using *in silico* bioinformatics algorithms for the accurate prediction of the impact of spike protein mutations on the pathogenicity, stability, and functionality of the SARS-CoV-2 virus and analysis of potential therapeutic targets. *Biochem Genet.* 2023;61(2):778-808. <https://doi.org/10.1007/s10528-022-10282-9>
27. Tan M, Xia J, Luo H, Meng G, Zhu Z. Applying the digital data and the bioinformatics tools in SARS-CoV-2 research. *Comput Struct Biotechnol J.* 2023;21:4697-4705. <https://doi.org/10.1016/j.csbj.2023.09.044>

28. Kirichenko AD, Poroshina AA, Sherbakov DY, Sadovsky MG, Krutovsky KV. Comparative analysis of alignment-free genome clustering and whole genome alignment-based phylogenomic relationship of coronaviruses. *PLoS One*. 2022;17(3):e0264640. <https://doi.org/10.1371/journal.pone.0264640>
29. van der Putten BC, Huijsmans NA, Mende DR, Schultsz C. Benchmarking the topological accuracy of bacterial phylogenomic workflows using *in silico* evolution. *Microb Genom*. 2022;8(3):000799. <https://doi.org/10.1099/mgen.0.000799>
30. Olawoye IB, Frost SD, Happi CT. The Bacteria Genome Pipeline (BAGEP): an automated, scalable workflow for bacteria genomes with Snakemake. *PeerJ*. 2020;8:e10121. <http://dx.doi.org/10.7717/peerj.10121>
31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>
32. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep*. 2021;11(1):12728. <https://doi.org/10.1038/s41598-021-91456-0>
33. Andersen PS, Stegger M, Aziz M, Contente-Cuomo T, Gibbons HS, Keim P, et al. Complete genome sequence of the epidemic and highly virulent CTX-M-15-producing H30-Rx subclone of *Escherichia coli* ST131. *Genome Announc*. 2013;1(6):e00988-13. <https://doi.org/10.1128/genomea.00988-13>
34. Campos-Madueno EI, Aldeia C, Perreten V, Sendi P, Moser AI, Endimiani A. Detection of blaCTX-M and blaDHA genes in stool samples of healthy people: comparison of culture- and shotgun metagenomic-based approaches. *Front Microbiol*. 2023;14:1236208. <https://doi.org/10.3389/fmicb.2023.1236208>
35. Khetrapal V, Mehershahi KS, Chen SL. Complete genome sequence of the original *Escherichia coli* isolate, strain NCTC86. *Genome Announc*. 2017;5(16):e00243-17. <https://doi.org/10.1128/genomea.00243-17>
36. Ramadan H, Soliman AM, Hiott LM, Elbediwi M, Woodley TA, Chattaway MA, et al. Emergence of multidrug-resistant *Escherichia coli* producing CTX-M, MCR-1, and FosA in retail food from Egypt. *Front Cell Infect Microbiol*. 2021;11:681588. <https://doi.org/10.3389/fcimb.2021.681588>
37. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. <https://doi.org/10.1093/bioinformatics/btt086>
38. Kikwai GK, Juma B, Nindo F, Ochieng C, Wamola N, Mbogo K, et al. Genome sequence of *Escherichia coli* isolated from an adult in Kibera, an urban informal settlement in Nairobi, Kenya. *Microbiol Resour Announc*. 2022;11(4):e01241-21. <https://doi.org/10.1128/mra.01241-21>
39. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647-4654. <https://doi.org/10.1093/molbev/msab199>
40. Ortega-Enríquez JZ, Arenas-Hernández MM, Barrios-Villa E. Draft genome sequence of a triple hybrid *Escherichia coli* strain isolated from a healthy donor feces. *Microbiol Resour Announc*. 2024;13(5):e00113-24. <https://doi.org/10.1128/mra.00113-24>
41. Aswal M, Singhal N, Kumar M. Genomic analysis of phylogroup D *Escherichia coli* strains using novel *de novo* reference-based guided assembly. *Sci Data*. 2023;10(1):573. <https://doi.org/10.1038/s41597-023-02444-0>

42. Coolen JP, den Drijver EP, Verweij JJ, Schildkraut JA, Neveling K, Melchers WJG, et al. Genome-wide analysis in *Escherichia coli* unravels a high level of genetic homoplasmy associated with cefotaxime resistance. *Microb Genom.* 2021;7(4):000556. <https://doi.org/10.1099/mgen.0.000556>
43. Lipworth S, Crook D, Walker AS, Peto T, Stoesser N. Exploring uncatalogued genetic variation in antimicrobial resistance gene families in *Escherichia coli*: an observational analysis. *Lancet Microbe.* 2024;5(11):e1009-e1018. [https://doi.org/10.1016/S2666-5247\(24\)00152-6](https://doi.org/10.1016/S2666-5247(24)00152-6)
44. Atlaw NA, Keelara S, Correa M, Foster D, Gebreyes W, Aidara-Kane A, et al. Identification of CTX-M-type ESBL-producing *Escherichia coli* from sheep and their abattoir environment using whole-genome sequencing. *Pathogens.* 2021;10(11):1480. <https://doi.org/10.3390/pathogens10111480>
45. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol.* 2012;50(4):1355-1361. <https://doi.org/10.1128/jcm.06094-11>
46. Nakamura A, Takahashi H, Arai M, Tsuchiya T, Wada S, Fujimoto Y, et al. Molecular subtyping for source tracking of *Escherichia coli* using core genome multilocus sequence typing at a food manufacturing plant. *PLoS One.* 2021;16(12):e0261352. <https://doi.org/10.1371/journal.pone.0261352>
47. Chowdhury PR, Hastak P, DeMaere M, Wyrsh E, Li D, Elankumaran P, et al. Phylogenomic analysis of a global collection of *Escherichia coli* ST38: evidence of interspecies and environmental transmission. *mSystems.* 2023;8(5):e01236-22. <https://doi.org/10.1128/msystems.01236-22>
48. Lees JA, Kendall M, Parkhill J, Colijn C, Bentley SD, Harris SR. Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation-based study. *Wellcome Open Res.* 2018;3:33. <http://dx.doi.org/10.12688/wellcomeopenres.14265.1>
49. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530-1534. <https://doi.org/10.1093/molbev/msaa015>
50. Wong TK, Ly-Trong N, Ren H, Demotte P, Baños H, Roger AJ, et al. IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Mol Biol Evol.* 2026;43(5):msag117. <https://doi.org/10.1093/molbev/msag117>
51. Cave R, Ter-Stepanyan MM, Mkrtchyan HV. Short- and long-read sequencing reveals the presence and evolution of an IncF plasmid harboring blaCTX-M-15 and blaCTX-M-27 genes in *Escherichia coli* ST131. *Microbiol Spectr.* 2023;11(4):e00356-23. <https://doi.org/10.1128/spectrum.00356-23>
52. Berwa A, Caspar Y. Easy analysis of bacterial whole-genome sequencing data for clinical microbiologists using open-source Galaxy platform: characterization of ESBL-producing Enterobacterales from bloodstream infections. *J Glob Antimicrob Resist.* 2024;39:153-158. <https://doi.org/10.1016/j.jgar.2024.08.012>
53. Peirano G, Chen L, Nobrega D, Finn TJ, Kreiswirth BN, DeVinney R, et al. Genomic epidemiology of global carbapenemase-producing *Escherichia coli*, 2015-2017. *Emerg Infect Dis.* 2022;28(5):924-932. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9045447/pdf/21-2535.pdf>
54. Maric L, Rupnik M, Janezic S. Diversity of ESBL-producing *Escherichia coli* in various water and sediment types. *PLoS One.* 2025;20(12):e0338703. <https://doi.org/10.1371/journal.pone.0338703>
55. Richter L, Duvenage S, Du Plessis EM, Msimango T, Dlangalala M, Mathavha MT, et al. Genomic evaluation of multidrug-resistant extended-spectrum β -lactamase (ESBL)-producing

- Escherichia coli* from irrigation water and fresh produce in South Africa: a cross-sectional analysis. *Environ Sci Technol.* 2024;58(32):14421-14438.
<https://doi.org/10.1021/acs.est.4c02431>
56. Fagerström A, Mölling P, Khan FA, Sundqvist M, Jass J, Söderquist B. Comparative distribution of extended-spectrum beta-lactamase-producing *Escherichia coli* from urine infections and environmental waters. *PLoS One.* 2019;14(11):e0224861.
<https://doi.org/10.1371/journal.pone.0224861>