Whole-Genome Sequencing, Annotation and Taxonomic Confirmation of a Multidrug-Resistant *Escherichia coli* Isolated from the Blood of a Sepsis Patient

Peter M. Eze[1,2*], Lucy Dillon[1], Jurnorain Gani[3], Timothy Planche[3], Linda Oyama[1], Chris Creevey[1]

[1]School of Biological Sciences, Queen's University Belfast, United Kingdom
[2]Department of Environmental Health Science, Nnamdi Azikiwe University, Nigeria
[3]Institute of Infection and Immunity, City St. George's, University of London, United Kingdom

*Corresponding author: p.eze@qub.ac.uk

ABSTRACT

Sepsis (blood stream infection) caused by multidrug-resistant (MDR) bacteria, particularly *Escherichia coli,* represents a significant global health threat due to high morbidity, mortality, and limited treatment options. *E. coli*, a major causative agent of bloodstream infections, has evolved highly virulent and MDR strains, which contribute to the increasing burden of antimicrobial resistance (AMR), complicating clinical management and reducing the efficacy of conventional antibiotic therapies. In this study, we characterised the genomic and phenotypic drug resistance mechanism of *E. coli* 266631E isolated from a sepsis patient, highlighting the negative implications of MDR *E. coli* in sepsis. Antimicrobial susceptibility testing and minimum inhibitory concentration analysis revealed resistance to multiple antibiotics, including amoxicillin, cefotaxime, ciprofloxacin, gentamicin, and tobramycin. Whole genome sequencing identified a broad array of AMR genes encoding resistance to various antibiotic classes, such as macrolides, fluoroquinolones, aminoglycosides, carbapenems, and cephalosporins. Notably, the CTX-M-15 gene, a key extended-spectrum β-lactamase determinant, was found in both the bacterial chromosome and an IncF-type plasmid, emphasizing the potential for horizontal gene transfer and rapid dissemination of resistance. Confirming the taxonomy of the novel and unidentified bacterial strain through querying its 16S rRNA sequence and genome in recognised bacterial taxonomic databases presented a challenge. The isolate showed genetic similarity to *E. coli*, *E. fergusonii*, and *Shigella* species despite their phenotypic differences and variations in their pathogenic traits. However, a simple phenotypic laboratory procedure, based on the biochemical and cultural differences among these bacteria in Coliform *ChromoSelect* Agar, confirmed the isolate as *E. coli*. This study highlights the importance of applying phenotypic approaches to support genomic identification of clinically significant bacteria, as this will help guide proper diagnosis and treatment of life-threatening infections like sepsis.

Keywords: *Escherichia coli*, sepsis, antimicrobial resistance, whole-genome sequence, phenotypic, genomic

INTRODUCTION

E. coli is an opportunistic pathogen but is typically a part of the gut microbiota of humans and other animals. It is one of the most versatile bacteria, capable of causing a variety of infections, including both intestinal and extraintestinal infections (Denamur et al., 2021; Priyanka et al., 2023). Extraintestinal pathogenic *E. coli* (ExPEC) comprises a group of strains with the ability to colonize and infect extraintestinal sites. Systemic infections caused by ExPEC are referred to as invasive *E. coli* disease (IED) or invasive ExPEC disease. This condition includes infections of the bloodstream and other normally sterile body sites, such as cerebrospinal fluid, the pleural cavity, peritoneal space, bone, and joints. Invasive *E. coli* disease may result in sepsis, septic shock, or death (Doua et al., 2023; Geurtsen et al., 2022; Rosenberg et al., 2021; Dugar et al., 2020). Colonization and infection of extraintestinal sites by pathogenic *E. coli* are leading causes of sepsis in both hospitals and communities (Markwart et al., 2020; Shao et al., 2023).

Escherichia coli is the most prevalent bacterium causing sepsis and one of the most common causes of bloodstream infections, as it is capable of accessing and surviving in the bloodstream (AL-Khikani et al., 2023; Shao et al., 2023). *E. coli* has been reported as the leading cause of bloodstream infections in high-income countries, with an estimated incidence of 48 cases per 100,000 person each year (Bonten et al., 2021; Maldonado et al., 2024). A study on early-onset neonatal sepsis involving 235 cases showed that the most frequent pathogen was *E. coli* (36.6%), which was associated with a higher incidence of mortality (Song et al., 2022; Stoll et al., 2020).

Sepsis has been defined as a dysregulated host immune response to infections, leading to a life-threatening organ dysfunction (Song et al., 2022). Sepsis is considered a major cause of mortality worldwide, being responsible for nearly 20% of global deaths (Maldonado et al., 2024).

Several multidrug-resistant (MDR) *E. coli* strains have been reported from sepsis patients (Kumar et al., 2024; Gashaw et al., 2024; Madrazo et al., 2023), making sepsis treatment challenging. Understanding the resistance mechanisms of sepsis-causing MDR pathogens enables the selection of effective antimicrobial agents, thereby improving clinical outcomes and minimising treatment failure. This approach also supports antimicrobial stewardship by reducing reliance on empiric antibiotic therapy and limiting the emergence of resistance.

This study describes the application of phenotypic and genomic approaches for the identification and antimicrobial characterization of a novel MDR *E. coli* strain isolated from the blood of a sepsis patient.

METHODS

Bacterial isolation and MALDI-TOF Identification

Following standard microbiological procedures for blood culture processing (CLSI, 2022) a bacterial isolate was obtained from blood sample of a sepsis patient a St. George's Hospital, London, United Kingdom. A fully automated microbial mass spectrometry detection system (Autof MS1600 MALDI-TOF, Autobio Labtec Instrument Co., Ltd., China) was used in the initial identification of the isolate. The equipment comprised of sample target plate, workstation, (microcomputer, LCD, capture card) instrument (matrix-assisted laser desorption/ionization time of flight detector, vertical ion flight tube, vacuum system), software and database. Sample preparation was carried out according to the Direct Coating Method described in the manual for MALDI-TOF analysis (available at https://chirus.com/autof-ms1600/). First, pure colonies of the bacterial isolates were aseptically transferred to the respective spots on the target plate. A volume of 1 µL of matrix solution (provided by the manufacturer) was then added to each spot containing the sample, and the target plate was placed on a heating block set at 40°C for 1 minute to dry. The target plate was then placed in a holder, which was inserted into the instrument and the analysis was performed. The generated mass spectra were automatically compared with the instrument's reference database of known spectra from different bacterial species. Based on the closest match from the database, the system provided an identification of the bacterial species.

Antibiotic Susceptibility Test (AST)

AST of the isolate was carried out using the disk diffusion method described by Chukwunwejim et al. (2018). A volume of 80 mL of molten MH agar (Sigma-Aldrich, India) was poured into 150 x 20 mm Petri plates and left to solidify. The isolate was grown overnight in MH broth. Using sterile swab stick, standardized broth culture (0.5 McFarland turbidity standard; OD600 ~ 0.08 - 0.1) of the isolate was swabbed onto the agar plates. Antibiotic discs [amoxicillin (25 µg), streptomycin (300 µg), cefotaxime (5 µg), oxytetracycline (30 µg), tobramycin (10 µg), and

ciprofloxacin (10 µg) (Oxoid, UK)] were placed on the agar surface at considerable distance between each other. The plates were left to stand for 15 min to allow for the antibiotics to diffuse, after which the plates were incubated at 37°C for 24 h. At the end of incubation, the plates were observed for IZDs, and these were measured in mm using a meter rule. *E. coli* (ATCC 25922), a non-MDR (antibiotic-susceptible) strain, was used as a control. The experiment was conducted in triplicate and the mean IZDs were calculated and recorded.

MIC Determination

The minimum inhibitory concentrations (MICs) of several antibiotics [ciprofloxacin (Merck, USA), gentamicin sulfate salt (Sigma-Aldrich, Germany), and neomycin sulphate (Apollo Scientific, UK)] on the isolate were determined by broth microdilution method in cation adjusted Mueller Hinton broth (CAMHB) (Sigma-Aldrich, India) using a final bacterial inoculum of $1x10^6$ CFU/ml (Oyama et al., 2022). The antibiotics, dissolved in sterile distilled water, were added to the bacterial culture and incubated overnight (18–24 h) at 37 °C. After incubation, the wells were visually examined for the presence or absence of growth, indicated by the turbidity of the medium. *E. coli* (ATCC 25922), a non-MDR strain, was used as a control. The MICs of the antibiotics were recorded as the lowest concentration that inhibited growth (prevented visible growth of bacteria).

Genomic DNA extraction, sequencing, and assembly

Genomic DNA extraction, hybrid (Illumina & Nanopore) sequencing, and assembly of the bacterial isolate was carried out by a commercial service provider, MicrobesNG (Birmingham, UK), according to protocols available at https://microbesng.com/documents/methods/. The bacterial DNA was subjected to HYBRID **(**Illumina & Nanopore) sequencing which combines data from the latest R10.4.1 chemistry from Oxford Nanopore Technologies with 2 x 250 bp Illumina reads to provide the absolute highest quality assemblies of bacterial genomes. Hybrid assembly was performed using Unicycler (version 0.4.0; Wick et al., 2017). QUAST on Galaxy https://usegalaxy.org/; Mikheenko et al. 2018; Gurevich et al., 2013) was used to assess the quality of the genome assembly.

Genome component prediction and gene annotation

The hybrid assembled genome was annotated using Prokaryotic Genome Annotation (Prokka) tool (version 1.14.6) available on Galaxy (https://usegalaxy.org/) (Seemann, 2014). PlasmidFinder (Galaxy Version 2.1.6+galaxy1) and MOB-Recon (Galaxy Version 3.0.3+galaxy0), both available on Galaxy, were also used to assess the properties of the contigs and determine whether they were chromosomes or plasmids, as well as to identify plasmid type and circularity (Carattoli et al., 2014; Robertson and Nash, 2018). "Extractseq" on Galaxy was used to extract the 16S rRNA sequence from the bacterial genome for species identification via BLAST searches on NCBI, Greengenes and SILVA databases. The GC content of the complete genome and the contigs were calculated using the GC Content Calculator available at https://jamiemcgowan.ie/bioinf/gc_content.html. Antimicrobial resistance genes (ARGs) were identified in the bacterial genome using tools available on Galaxy: AMRFinderPlus (Galaxy Version 3.12.8+galaxy0), an NCBI antimicrobial resistance gene (ARG) finder, and starAMR (Galaxy Version 0.11.0+galaxy0), which searched against the ResFinder, PlasmidFinder, and PointFinder databases. Additionally, the Resistance Gene Identifier (RGI) from the Comprehensive Antibiotic Resistance Database (CARD), available at https://card.mcmaster.ca/analyze/rgi, was used to identify the ARGs in the bacterial genome. Functional annotation of the bacterial genome was performed using the eggNOG-mapper database, available at http://eggnog-mapper.embl.de/ (Cantalapiedra et al., 2021).

Isolate identity confirmation by 16S taxonomy

To confirm the identity of the isolate, its 16S rRNA gene (in FASTA format) was analysed using the standard and rRNA/ITS databases of the NCBI BLAST search tool, available at https://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschul et al., 1990). Additionally, the 16S rRNA sequence was queried against the SILVA database (https://www.arb-silva.de/) using the SILVA ACT (Alignment, Classification, and Tree) tool, which allows for sequence upload and comparison against their curated ribosomal RNA database (Pruesse et al., 2012). The appropriate search parameters, such as targeting the 16S rRNA gene for bacteria, were selected, and the search was executed. The 16S rRNA sequence was also analysed using the Greengenes2 platform (McDonald et al., 2024), which is designed to classify microbial 16S rRNA sequences based on taxonomic hierarchy, comparing the sequence against a curated 16S rRNA database to identify close matches and provide taxonomic assignments.

Sourmash (version 4.8.10) was used to cluster the isolate with closely related species identified in the 16S rRNA NCBI BLAST search, aiming to clarify its taxonomic species-specific identity by comparing it to various closely related species' genomes found in the BLAST search (Brown and Irber, 2016). To avoid misclassification due to plasmid presence, only the bacterial chromosome was analysed. A total of 33 additional genomes of closely related species, identified in the 16S rRNA BLAST search, were downloaded from the NCBI database, with two genomes for each bacterial species identified as close hits in the BLAST search. These genomes, along with the isolate's chromosome were analysed in Sourmash using the Jaccard Index with a k-mer size of 31. To further confirm the results, a second Sourmash plot was generated using the isolate's chromosome and an additional 1,594 bacterial genomes downloaded from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) database, available at https://www.bv-brc.org/ (see *wget.sh* and *genome_list.txt* at https://github.com/LucyDillon/MDR_isolate for details). The estimated average nucleotide identity (ANI) and Jaccard Index of the 1,595 isolates were compared using a k-mer size of 31.

Phylogenetic relatedness of isolate

A pangenome analysis was conducted to produce a core genome alignment of the closely related species based on the 16S rRNA BLAST search. To generate the core gene alignment, the genomes were first annotated using Prokka (version 1.14.5; Seemann, 2014). Due to the large number of genomes, this step was performed via high-performance computing (HPC). Roary (version 3.13.0) was then used for the pangenome analysis (Page et al., 2015) (see *Roary_core gene.sh* at https://github.com/LucyDillon/MDR_isolate for parameter details). Roary produces a default Newick file via FastTree (Price et al., 2009), which was visualized using iTOL (Letunic and Bork, 2024). Additionally, another phylogenetic tree was generated from the core gene alignment using IQ-TREE (version 2; Minh et al., 2020), and the resulting tree was visualized in iTOL (Nguyen et al., 2015) (see *iq_tree.sh* at https://github.com/LucyDillon/MDR_isolate for parameter details).

Biochemical differentiation and phenotypic confirmation of isolate

To confirm the phenotypic and biochemical similarity or dissimilarity of the isolate to *E. coli* or *E. fergusonii,* and to ascertain if the isolate is a *Shigella* species, the isolate and reference strains of *E. coli* (ATCC 25922) and *E. fergusonii* (ATCC 35469) were streaked on Coliform *ChromoSelect* Agar (CCA; Sigma-Aldrich, Germany) and incubated overnight at 37°C. CCA is a selective medium used for the detection of *E. coli* and total coliforms in water and food samples. The chromogenic mixture contains two chromogenic substrates: Salmon-GAL and X-glucuronide. The enzyme β-D-galactosidase, produced by coliforms (including *E. fergusonii* strains), cleaves Salmon-GAL, resulting in a colour change from salmon to red in the coliform colonies. *E. coli* produces dark blue to violet colonies on CCA because it cleaves both Salmon-GAL and X-glucuronide using β-D-galactosidase and β-D-glucuronidase, respectively, and exhibits a positive indole reaction, as the medium contains tryptophan. *Shigella* spp., being non-coliform, appear colourless on CCA, as they are negative for Salmon-GAL, X-Gluguronide, and indole (Sigma-Aldrich, 2024).

Ethical approval

Approval for this study was received from the Research Ethics Committee of Queen's University Belfast's School of Biological Sciences.

RESULTS

Preliminary Strain identification and Phenotypic Resistance Profiling

MALDI-TOF analysis identified the isolate as *Escherichia coli* (see Supplementary Material: Figure S1). It was then ascribed the strain-level identity *E. coli* 266631E. Resistance to several commonly used antibiotics was confirmed through disk diffusion testing and MIC determination (Table 1). Compared to the non-MDR reference strain *E. coli* strain (ATCC 25922), the isolate exhibited resistance to amoxicillin, cefotaxime, tobramycin and ciprofloxacin, with inhibition zone diameters (IZD) of 0 mm for all except for tobramycin, which had an IZD of 11 mm. Additionally, MIC determination by broth dilution assay revealed that the isolate was highly resistant to ciprofloxacin and gentamicin, with MIC values of 31.25 and 500 µg/mL, respectively, compared to the non-MDR control strain.

Table 1: Phenotypic Resistance Profile of the Bacterial Isolate

Antibiotics	**Inhibition Zone Diameter (mm)**	
	E. coli 266631E	*E. coli* (ATCC 25922)
Amoxicillin (25 µg)	0	20
Streptomycin (300 µg)	22	22
Cefotaxime (5 µg)	0	28
Oxytetracycline (30 µg)	23	26
Tobramycin (10 µg)	11	20
Ciprofloxacin (10 µg)	0	38
	MIC (µg/mL)	
Ciprofloxacin	31.25	0.015
Gentamicin	500	3.9
Neomycin	3.9	3.9

Genomic Insights into Antimicrobial Resistance and Functional Gene Annotations

The complete genome sequence of *E. coli* 266631E is 5,454,694 bp in size and consists of 10 contigs. Contig 1 represents the chromosome (5,019,723 bp), while the remaining nine contigs include those containing plasmids such as IncF-type (contigs 3 and 4), Col156 (contigs 5), and ColMG828 (contig 8) (Table 2; Supplementary Material: Figures S2 – S5). ARGs identified using AMRFinderPlus, starAMR and RGI revealed a collection of ARGs including *aadA5, CTX-M-15, dfrA17, mphA, qacE, sul1,* and *TEM-1* as outlined in Table 2 below.

Using eggNOG-mapper, the functional gene families across the bacterial isolate were identified. The COG (clusters of orthologous groups) categories that constituted the largest gene families within the bacterial genome were M (Cell wall/membrane/envelope biogenesis), P (Inorganic ion transport and metabolism), and K (Transcription) (Figure 1). The majority of the RGI predicted ARGs identified in the bacterial chromosome were grouped into their respective COG categories, as indicated in the eggNOG-mapper output file (Table 3).

Table 2: Genome assembly information and ARG carriage in *E. coli* 266631E

Contig	Total length (bp)	GC content	Plasmid type [As determined by starAMR (†), Plasmid finder (‡), and MOB-Recon (§)]	Circularity status [As determined by Unicycler (†), Plasmid finder (‡), and MOB-Recon (§)]	Predicted ARGs [As determined by AMRFinderPlus (%Identity: 100%)] and starAMR (%Identity: ≥99%)]	Predicted ARGs [As determined by RGI (%Identity: ≥99%)]	NCBI Genome Accession Number*
Contig 1 (Chromosome)	5019723	50.77%	-	-	*CTX-M-15* (2 copies)	*acrA, AcrAB-TolC* (with *AcrR* mutation), *acrB, acrD, AcrE, AcrF, bacA, cpxA, CRP, CTX-M-15* (2 copies), *emrA, emrB, emrR, emrY, evgA, gadW, H-NS, kdpE, leuO, marA, mdtB, mdtE, mdtF, mdtG, mdtH, msbA, PmrF, soxR, soxS, TolC, ugd, YojI*	CP183489
Contig 2	218237	50.13%	-	Circular[†§]	None	None	CP183490
Contig 3	126566	52.54%	IncF-type[†‡§]	Circular[†‡§]	*aac(3)-II, aadA5, CTX-M-15, dfrA17, mph(A), qacE, sul1, TEM-1B*	*aadA5, CTX-M-15, dfrA17, mphA, Mrx, qacEdelta1, sul1, TEM-1*	CP183491
Contig 4	74099	51.25%	IncF-type[†‡§]	Circular[†‡§]	None	None	CP183492
Contig 5	6078	49.03%	Col156[‡§]	Circular[†‡§]	None	None	CP183493
Contig 6	4600	45.74%	-	-	None	None	CP183494
Contig 7	1709	54.13%	-	-	None	None	CP183495
Contig 8	1552	51.87%	ColMG828[§]	Circular[†§]	None	None	CP183496
Contig 9	1307	47.28%	-	Circular[†§]	None	None	CP183497
Contig 10	823	50.67%	-	-	None	None	CP183498

*The isolate's complete genome, including the chromosome, plasmids, and the other contigs, was deposited in the NCBI Genome database under the accession numbers CP183489 – CP183498.

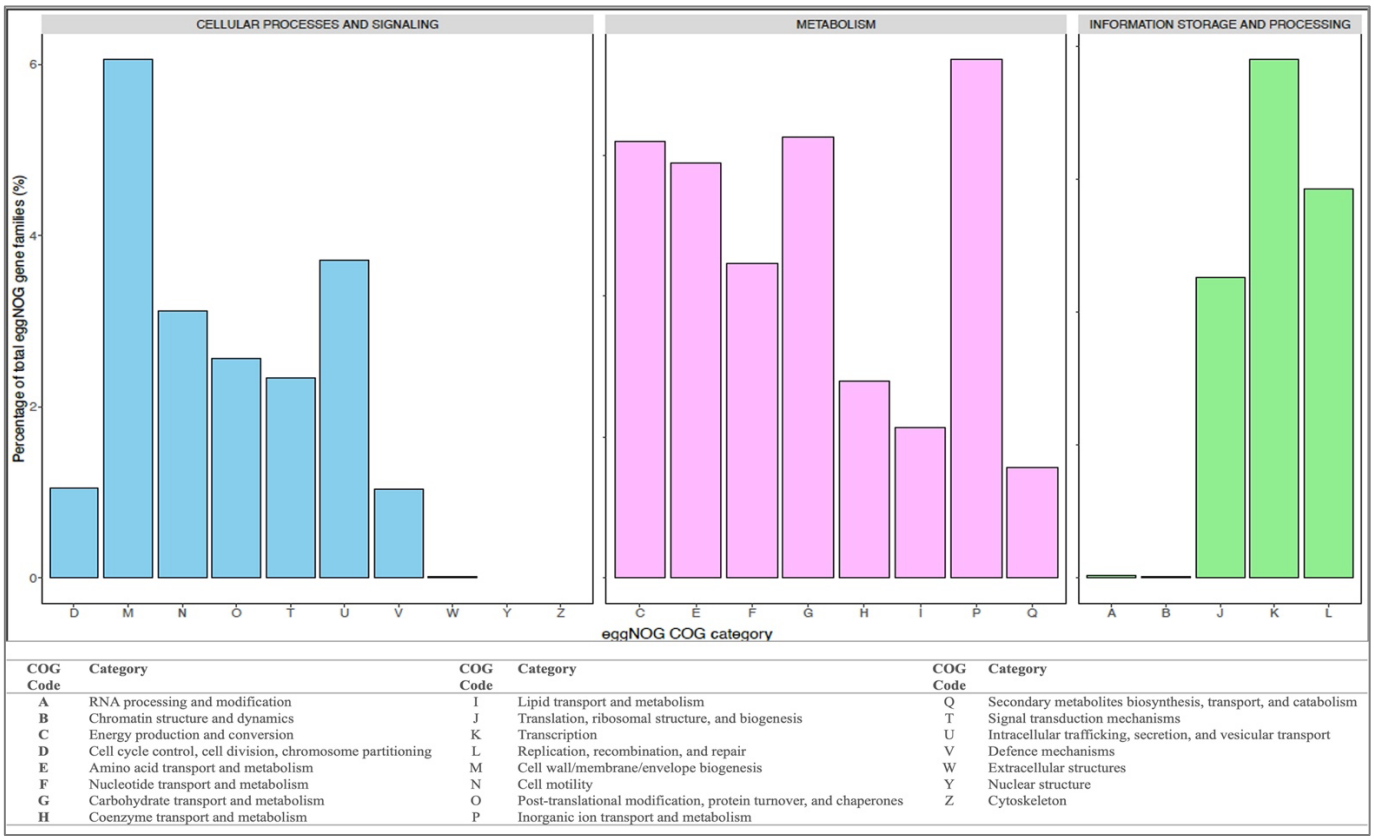

Figure 1: The distribution of eggNOG gene families across *E. coli* 266631E genome

Table 3: Classification of the identified ARGs into functional gene families by COG categories

COG category	RGI predicted genes		
	ARGs	AMR mechanism	Target Drug Class
M (Cell wall/membrane/envelope biogenesis)	*acrA*	Antibiotic target alteration, antibiotic efflux	Fluoroquinolones, cephalosporins, glycylcyclinse, penams, tetracyclines, rifamycins, phenicols, disinfecting agents, antiseptics
	tolC	Antibiotic efflux	Macrolides, fluoroquinolones, aminoglycosides, carbapenem, cephalosporins, glycylcycline, cephamycins, penams, tetracyclines, peptide antibiotic, aminocoumarins, rifamycins, phenicols, penems, disinfecting agents and antiseptics
	mdtE	Antibiotic efflux	Macrolides, fluoroquinolones, penam
P (Inorganic ion transport and metabolism)	*yojI*	Antibiotic efflux	Peptide antibiotic
K (Transcription)	*soxS*	Antibiotic target alteration, antibiotic efflux, reduced permeability to antibiotic	Fluoroquinolones, monobactams, carbapenems, cephalosporins, glycylcyclines, cephamycins, penam, tetracyclines, rifamycins, phenicols, penems, disinfecting agents, antiseptics
	soxR	Antibiotic target alteration, antibiotic efflux	Fluoroquinolones, cephalosporins, glycylcyclines, penams, tetracyclines, rifamycins, phenicols, disinfecting agents and antiseptics
	crp	Antibiotic efflux	Macrolides, fluoroquinolones, penams
	evgA	Antibiotic efflux	Macrolides, fluoroquinolones, penams, tetracyclines
	kdpE	Antibiotic efflux	Aminoglycosides
	H-NS	Antibiotic efflux	Macrolides, fluoroquinolones, cephalosporins, cephamycins, penams, tetracyclines
	gadW	Antibiotic efflux	Macrolides, fluoroquinolones, penam

	AcrAB-TolC (with *AcrR* mutation)	Antibiotic target alteration, antibiotic efflux	Fluoroquinolones, cephalosporins, glycylcyclinse, penams, tetracyclines, rifamycins, phenicols, disinfecting agents, antiseptics
	marA	Antibiotic efflux, reduced permeability to antibiotics	Peptide antibiotics
	leuO	Antibiotic efflux	Nucleosides, disinfecting agents, antiseptics
U (Intracellular trafficking, secretion, and vesicular transport)	*acrD*	Antibiotic efflux	Aminoglycosides
	emrB	Antibiotic efflux	Fluoroquinolones
V (Defence mechanisms)	*acrB*	Antibiotic efflux	Fluoroquinolones, cephalosporins, glycylcyclines, penams, tetracyclines, rifamycin, phenicol, disinfecting agents, antiseptics
	mdtB	Antibiotic efflux	Aminocoumarins
	mdtF	Antibiotic efflux	Macrolides, fluoroquinolones, penams
	msbA	Antibiotic efflux	Nitroimidazoles
	emrA	Antibiotic efflux	Fluoroquinolones
T (Signal transduction mechanisms)	*cpxA*	Antibiotic efflux	Aminoglycosides, aminocoumarins
C (Energy production and conversion)	*ugD*	Antibiotic target alteration	Peptide antibiotics

Taxonomy

- ***Query of the isolate's 16S rRNA sequence against NCBI, Greengenes2, and SILVA Databases***

A BLAST search of the bacterial 16S rRNA sequence in the NCBI standard databases identified the isolate as *E. coli* with 100% identity (Supplementary Material: Figure S6). However, a BLAST search of the bacterial 16S rRNA sequence in the NCBI rRNA/ITS databases showed the highest percent identity to two *Shigella* species – *S. sonnei* (99.80%) and *S. flexneri* (99.80%), followed by two *E. fergusonii* strains (99.66% and 99.54%, respectively) and then *E. coli* (99.52%) (Supplementary Material: Figure S7). Furthermore, querying the bacterial 16S rRNA sequence against the Greengenes2 database revealed 100% identity matches for 1,487 out of 6,066 bacterial strains in the database. Among these 1,487 strains, the highest species-specific match was for *E. fergusonii* with 130 matches, followed by 30 matches for *E. coli*, and 10 matches for *Shigella* species. Interestingly, most of the 100% matches were for non-species-specific bacteria in the *Escherichia* genus (Supplementary Material: Figure S8). Finally, a sequence alignment of the bacterial 16S rRNA sequence against the SILVA reference database yielded a non-species-specific and non-genus-specific taxonomic classification: 'Escherichia-Shigella' (Supplementary Material: Figure S9).

- ***Phylogeny of the bacterial isolate***

As the query of the isolate's 16S rRNA sequence against NCBI, Greengenes2, and SILVA databases did not align in terms of taxonomy, Sourmash was used to cluster the isolate (labelled as "unknown isolate"; Figure 2) with additional genomes. A total of 33 additional genomes of closely related species, identified in the 16S rRNA NCBI BLAST search, were downloaded, with two genomes for each bacterial species identified as close hits in the BLAST search (i.e., *E. coli*, *E. fergusonii* and *Shigella* spp.). Sourmash Jaccard index analysis showed that the closest genome was *E. coli*; however, several *Shigella* spp. clustered closer to the isolate than the second *E. coli* genome, making it difficult to definitively identify the isolate as *E. coli*.

Since the initial Sourmash analysis was not entirely conclusive—possibly due to the small number of genomes analysed—and to enable a more comprehensive Sourmash clustering analysis, an additional 1,594 genomes were sourced from the BV-BRC database (a comprehensive, curated repository of bacterial and viral genomes that provides access to a vast collection of high-quality, annotated reference sequences). Both the ANI plot (Supplementary Material: Figure S10) and the Jaccard Sourmash plot (Supplementary Material: Figure S11) showed that the isolate clustered among several *Shigella* and *Escherichia* species.

Due to the ambiguous results of the Sourmash analysis regarding the species-specific identification of the isolate, a pangenome analysis was conducted to produce a core genome alignment of closely related species identified by the 16S rRNA BLAST search. The 1,595 isolates from the Sourmash analysis were used to construct the phylogeny tree.

Roary was employed to generate a core gene alignment and create a Newick file with FastTree, which was visualized in iTOL (Supplementary Material: Figure S12). The isolate clustered with a mix of *Shigella* and *Escherichia* species, consistent with the patterns observed in the Sourmash plots.

The phylogenetic tree produced by FastTree did not provide a definitive or conclusive taxonomy (i.e., species-specific identification) for the isolate. Therefore, a more rigorous phylogenetic tool, IQ-TREE, was used to generate another tree from the core gene alignment. IQ-TREE is known for its advanced modelling of evolutionary processes, stronger statistical support for tree branches, and greater reliability for species-level resolution—particularly in closely related taxa such as *Escherichia* and *Shigella* (Nguyen et al., 2015; The et al., 2021; Minh et al., 2020). The resulting tree was visualized in iTOL, and branches from unrelated clades were removed to provide a clearer visualization of the isolate and surrounding bacterial species. The IQ-TREE plot yielded the same result as the FastTree plot, as well as the Sourmash plots, showing the isolate clustering with *Escherichia* and *Shigella* species (Figure 3).

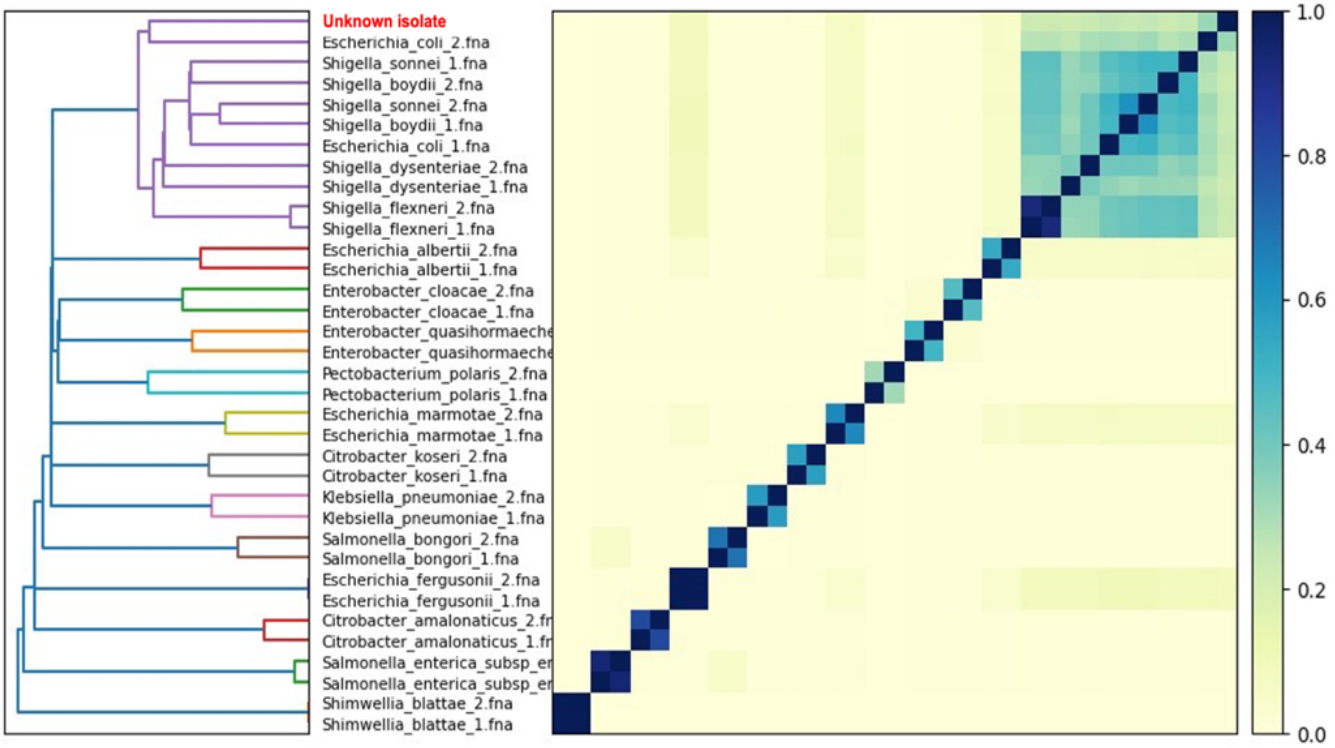

Figure 2: Sourmash plot of 34 genomes (the isolate's chromosome plus 33 additional genomes of closely related species identified in the 16S rRNA BLAST search). The plot showed that the closest genome to the isolate is *E. coli*. However, several *Shigella* spp. also clustered closer to the isolate than the second *E. coli* genome, complicating the definitive identification of the isolate as *E. coli*.

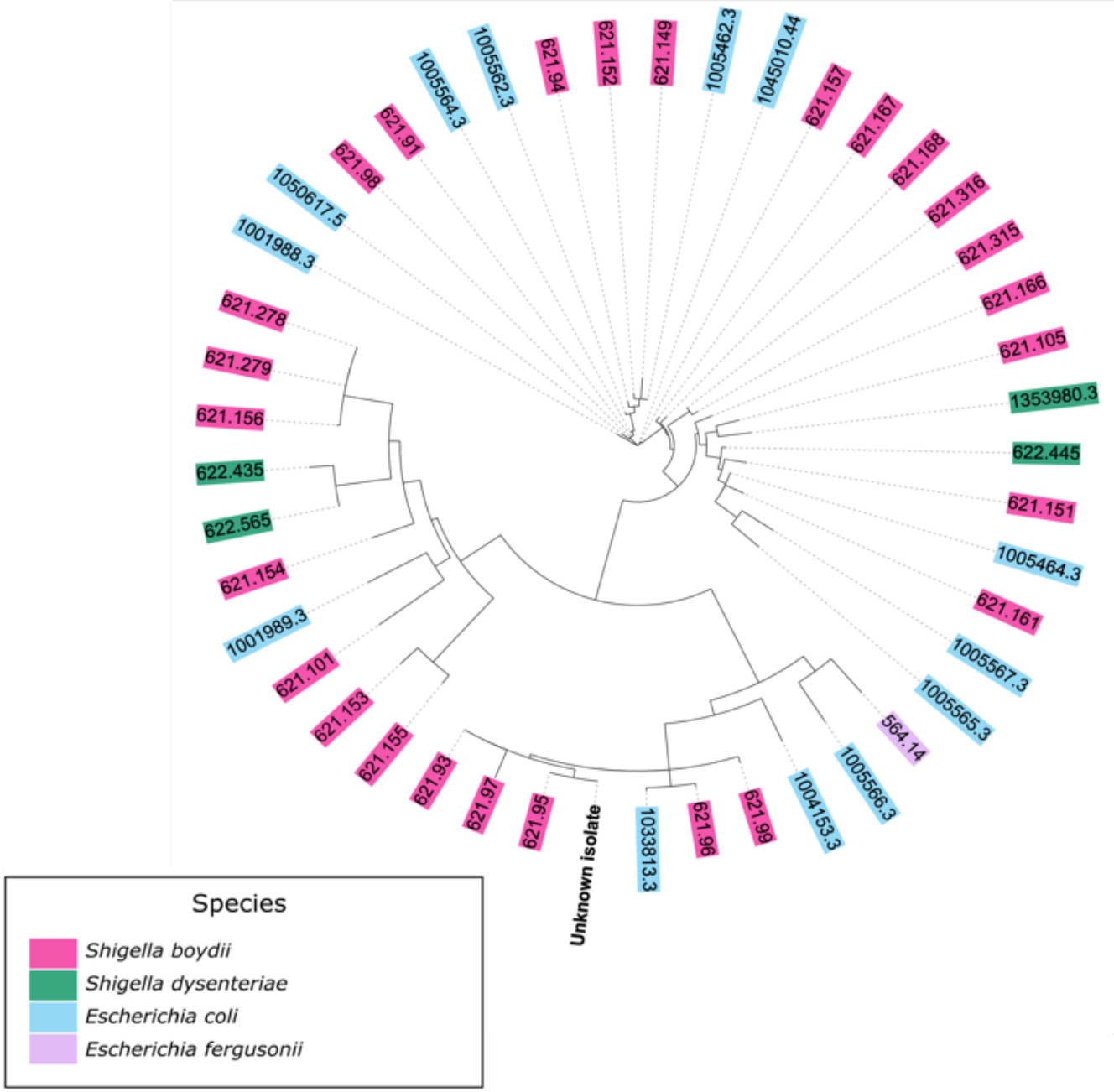

Figure 3: Phylogeny of the isolate using IQ-TREE. The tree is a reduced representation of 1,595 isolates; and clades have been removed to be able to visualise the isolate (labelled as "unknown isolate") and surrounding genomes.

- ***Biochemical differentiation of isolate***

By applying a simple phenotypic, biochemical, and cultural approach to confirm the identity of the isolate or distinguish it from *E. coli*, *E. fergusonii*, or *Shigella* species, the isolate and typed strains of *E. coli* (ATCC 25922) and *E. fergusonii* (ATCC 35469) were cultivated on CCA, a chromogenic selective agar that differentiates *E. coli* from other coliforms. The isolate, like the typed *E. coli* strain (ATCC 25922), produced characteristic dark blue to violet-coloured colonies on CCA due to cleavage of both Salmon-GAL and X-glucuronide in the selective chromogenic culture medium. In contrast, *E. fergusonii* (ATCC 35469) produced pink colonies (see Figure 4). Based on the biochemical/chromogenic characteristics, the isolate can be identified as an *E. coli* strain and is clearly distinguishable from *E. fergusonii,* and from *Shigella* species, which are known to produce colourless colonies on CCA (Sigma-Aldrich, 2024).

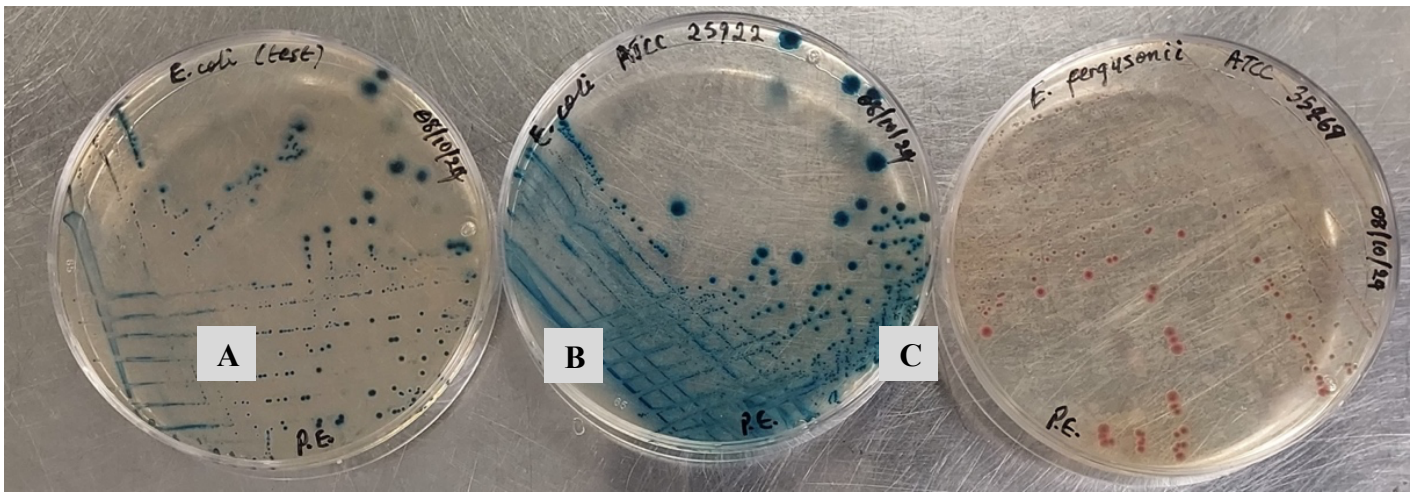

Figure 4: The colony characteristics of the isolate (A) on CCA showed dark blue to violet pigmentation, typical of *E. coli*, as confirmed by the dark blue to violet colonies of the reference *E. coli* strain (ATCC 25922) (B). The isolate is clearly distinguishable from the pink colonies of *E. fergusonii* (C) and from *Shigella* species, which are known to produce colourless colonies on CCA.

DISCUSSION

E. coli, one of the most extensively studied microorganisms in the world, is an opportunistic pathogen comprising strains that typically exist as harmless bacteria, forming a normal part of the gastrointestinal flora in many warm-blooded animals and reptiles. However, several highly adapted commensal clones can acquire specific virulence attributes, becoming highly pathogenic and capable of causing life-threatening infections, such as sepsis (Peng et al., 2024; Geurtsen et al., 2022).

E. coli sepsis, an invasive ExPEC disease, is a systemic infectious disease characterized by severe fever, chills, tachycardia, hypotension, and multiple organ dysfunction syndrome, which are typical clinical symptoms (Shao et al., 2023; Doua et al., 2023). ExPEC is now the leading cause of invasive bacterial disease globally, surpassing pathogens like *Staphylococcus aureus* and *Streptococcus pneumoniae*. *E. coli* is responsible for over 40% of bloodstream infections, with a case-fatality rate (CFR) of 12.4%. Sepsis, often caused by *E. coli,* accounted for 11 million deaths in 2017 from nearly 49 million cases globally (Doua et al., 2023; Pérez-Crespo et al., 2021; Bonten et al., 2021; Buetti et al., 2016).

Early sepsis diagnosis and timely and appropriate antibiotic therapy are crucial in preventing the progression of sepsis to septic shock or multiorgan failure. However, treatment can be delayed due to misdiagnosis or lack of knowledge about the pathogen (Doua et al., 2023). Antimicrobial resistance in ExPEC, particularly concerning MDR strains that cause sepsis, such as extended-spectrum β-lactamase (ESBL)-producing *E. coli*, poses a major threat to effective treatment. This underscores the need to understand resistance mechanisms to guide antibiotic selection and improve survival rates (Kumar et al., 2024; Gashaw et al., 2024; Doua et al., 2023; Madrazo et al., 2023).

This study reports the isolation and identification of an MDR *E. coli* strain (*E. coli* 266631E) from the blood of a sepsis patient, as well as the characterization of the genetic determinants of AMR in the bacterial isolate.

Phenotypic AMR Screening and Antimicrobial Resistance Gene (ARG) Analysis

Both phenotypic AMR screening (disk diffusion testing and broth microdilution MIC determination) and ARG analysis demonstrated the multi-drug resistance of the bacterial isolate. Results from the AST and MIC determination experiments show that *E. coli* 266631E was resistant to multiple antibiotics, including amoxicillin, cefotaxime, tobramycin, ciprofloxacin, and gentamicin (Table 1). WGS and analysis of the genetic determinants of AMR in the isolate reveal a multitude of ARGs responsible for diverse mechanisms of resistance to a vast range of antimicrobials across different classes.

Genomic analysis of *E. coli* 266631E revealed genes encoding for resistance to clinically important antibiotics, including those from the macrolide, fluoroquinolone, aminoglycoside, cephamycin, glycylcycline, tetracycline, peptide antibiotic, aminocoumarin, rifamycin, phenicol, cephalosporin, penam, and penem classes (Tables 2 and 3).

Multiple ARGs with diverse mechanisms of action and responsible for resistance to a wide range of antibiotics from different classes, were detected in the bacterial chromosome using RGI (Table 2). However, only *CTX-M-15* was identified in the chromosome using the AMRFinderPlus and starAMR databases. Compared to AMRFinderPlus and starAMR, RGI appears to be more robust in identifying ARGs in the bacterial chromosome. Outside the chromosome, ARGs were detected only in contig 3, identified as an IncF-type plasmid. In this plasmid, except for the *Mrx* gene, which was detected only by RGI, and *aac(3)-II*, detected only by AMRFinderPlus and starAMR, all three AMR prediction tools (AMRFinderPlus, starAMR, and RGI) equally detected seven similar ARGs: *aadA5, CTX-M-15, dfrA17, mphA, qacE, sul1,* and *TEM-1.*

The *CTX-M-15*, a key ARG of public health concern and one of the most common and influential genotypes of the $bla_{\text{CTX-Ms}}$ family, is widely present in *E. coli* (Yu et al., 2024; Rangama et al., 2021). This gene, identified both in the bacterial chromosome (with duplicate copies in different loci) and the IncF-type plasmid identified in contig 3 (Table 2), is known to encode for ESBL, which targets beta-lactam antibiotics, including third-generation cephalosporins (such as cefepime, cefotaxime, ceftriaxone) and penicillins (Rossolini et al., 2008; Livermore and Hawkey, 2005). Resistance to carbapenems has also been observed in *CTX-M-15*-producing bacteria (Elliott et al., 2006; Livermore and Hawkey, 2005). As observed in the *E. coli* strain isolated in this study, this MDR gene is frequently carried on plasmids, facilitating its rapid spread between different bacterial species (Yu et al., 2024; Grevskott et al., 2024).

It is also interesting to note the high level of quinolone resistance exhibited by *E. coli* 266631E. AST results show that the isolate is highly resistant to ciprofloxacin, with a disc concentration of 10 µg producing an IZD of 0 mm, and a MIC of 31.25 µg/ml (Table 1). In contrast, the susceptible reference strain *E. coli* (ATCC 25922) displayed an IZD of 38 mm and an MIC of 0.015 µg/ml, which falls within EUCAST's MIC range of 0.004–0.016 µg/ml for ciprofloxacin against *E. coli* (ATCC 25922) (EUCAST, 2023). Quinolone resistance is confirmed by the presence of multiple quinolone-resistance-encoding ARGs found in the bacterial genome. A majority of the ARGs identified in the bacterial chromosome are associated with fluoroquinolone resistance. These genes include: *acrA, acrB, crp, evgA, gadW, H-NS, AcrAB-TolC* (with *AcrR* mutation), *emrA, emrB, mdtE, mdtF, soxR, soxS,* and *tolC.* These ARGs are able to induce resistance to quinolones through diverse mechanisms including antibiotic efflux, antibiotic target alteration and reduced permeability to antibiotics. No quinolone-resistance-encoding ARGs were found in the IncF-type plasmid identified in contig 3 (Table 2; Supplementary Material: Figures S2 – S5).

An MIC of 500 µg/mL against gentamicin, along with reduced susceptibility to tobramycin at a disc concentration of 10 µg (11 mm IZD) by the isolate, compared to the reference strain *E. coli* (ATCC 25922), which exhibited an MIC of 3.9 µg/mL against gentamicin and a 20 mm IZD to tobramycin (Table 1), indicates a high level of resistance to aminoglycosides. Several ARGs responsible for aminoglycoside resistance were found in the bacterial chromosome (*tolC, kdpE, acrD, cpxA*), as well as in the IncF-type plasmid identified in contig 3 (*aadA5* and *aac(3)-II*) (Table 2; Supplementary Material: Figures S2 – S5).

The majority of the ARGs identified in the bacterial chromosomes were categorized into functional gene families using eggNOG-mapper, with most falling into COG categories M, K, and V (Table 3, Figure 1). Table 3 shows that ARGs within each COG category exhibit diverse resistance mechanisms. Additionally, some genes with similar resistance mechanisms were categorized into different COG categories. For example, most ARGs confer resistance through the antibiotic efflux mechanism, and these genes are distributed across various COG categories, including M, P, K, U, V, and T. From the eggNOG-mapper functional gene family categorization (Figure 1), it can be inferred that ARGs classified within the three highly expressed COG categories – M (cell wall/membrane/envelope biogenesis), P (inorganic ion transport and metabolism), and K (transcription) play a significant role in the MDR characteristics of the isolate.

Taxonomy

An attempt to confirm the identity of the bacterial isolate by querying the bacterial 16S rRNA sequence against the NCBI, Greengenes, and SILVA databases produced variable outcomes that did not align taxonomically. A BLAST search of the bacterial 16S rRNA sequence in the NCBI standard databases identified the isolate as *E. coli* with 100% identity. However, when searched in the NCBI rRNA/ITS databases, the sequence showed the highest identity to *Shigella* species, followed by *E. fergusonii,* and then *E. coli* (Supplementary Material: Figures S6 – S7). Querying the bacterial 16S rRNA sequence against the Greengenes2 database revealed 100% identity matches for *E. fergusonii*, *E. coli*, and *Shigella* species. Additionally, submitting the 16S rRNA sequence to SILVA yielded a non-

species-specific and non-genus-specific taxonomic classification: '*Escherichia–Shigella*' which is consistent with the report by Edgar (2018). Edgar noted that SILVA, which is based on Bergey's Manual and List of Prokaryotic Names with Standing in Nomenclature (LSPN), addresses the taxonomic overlap between *Escherichia* and *Shigella* by retaining well-established species names, such as *Escherichia coli,* and introducing a combined genus named *Escherichia–Shigella*.

The taxonomic relationship between *Escherichia* and *Shigella* has been a longstanding subject of debate, mainly due to the close genetic and phenotypic similarities between these genera (Sanford et al., 2021; Khot and Fisher, 2013; Lan and Reeves, 2001). Traditional taxonomic methods, such as 16S rRNA sequencing, which has been widely used to classify bacterial species, lack the resolution to effectively distinguish between closely related species such as *Escherichia* and *Shigella* (Lan and Reeves, 2001; Rantakokko-Jalava et al., 2000). The 16S rRNA gene is conserved across most bacteria, making it useful for broader taxonomic classifications (Jo et al., 2017) but insufficient for species-level distinctions in many genera, including *Escherichia* and *Shigella* (Liu et al., 2023; Khot and Fisher, 2013).

Historically, *Shigella* was classified separately from *E. coli* based on clinical, phenotypic, and biochemical differences, especially due to its association with dysentery and other pathogenic traits. There are very few biochemical properties that distinguish *Shigella* from enteroinvasive *E. coli* (EIEC), which is also a major cause of dysentery. In fact, some O-antigens associated with EIEC are identical to those found in *Shigella* spp., and many plasmid-associated virulence determinants are common to both EIEC and *Shigella* (Yang et al., 2005; Lan et al., 2001; Lan and Reeves, 2001; Cheasty and Rowe, 1983). However, advances in genomics, particularly whole genome sequencing, have shown that *Shigella* is genetically nested within the *Escherichia* genus, raising questions about the validity of the historical separation of these genera (Abram et al., 2021; The et al., 2016; Yang et al., 2005). Because they are genetically almost identical to certain strains of *E. coli* but have distinct pathogenic mechanisms that cause specific diseases, particularly dysentery (shigellosis), *Shigella* spp. are considered to be pathotypes of *E. coli* (Pupo et al., 2000, Lan et al., 2004; Lindsey et al., 2017). Despite genomic evidence supporting the integration of *Shigella* into *Escherichia*, the practical implications of such a taxonomic change are significant and could impact clinical diagnostics, epidemiological tracking, and communication among researchers. Therefore, any taxonomic changes must be carefully considered and widely accepted by both the scientific and medical communities to avoid confusion and ensure clarity in bacterial classification.

The sequencing of the 16S rRNA gene is commonly utilized to survey microbial communities, with specialized databases like Greengenes, RDP, and SILVA developed to support this method. It is recognized that most taxonomic annotations in these databases are predictions derived from sequences, rather than authoritative classifications based on studies of type strains or isolates (Edgar, 2018). Furthermore, it is acknowledged that databases like Greengenes and SILVA contain a considerable number of taxonomies that lack species-level resolution, which has limited the performance of classifiers (Hsieh et al., 2022).

The shortcomings of 16S rRNA sequencing extend beyond *Escherichia* and *Shigella*. Many bacterial genera, such as *Pseudomonas*, *Aeromonas*, and *Streptococcus*, exhibit similar problems where 16S rRNA-based classifications lack the discriminatory power to delineate species accurately (Jenkins et al., 2012; Janda and Abbott, 2007). The broader implications of this taxonomic debate highlight the need for standardized frameworks in bacterial classification. As 16S rRNA databases like Greengenes, SILVA, and RDP continue to evolve, taxonomic inconsistencies, particularly in closely related species like *Shigella* and *E. coli*, pose ongoing challenges. Therefore, a unified nomenclature system is essential for ensuring consistency and interoperability in bacterial taxonomy (Edgar, 2018).

Since querying the isolate's 16S rRNA sequence and genome in recognised bacterial taxonomic databases did not yield a definitive or consistent taxonomic classification, Sourmash was used to cluster the isolate with additional genomes of bacterial species identified as close hits in the 16S rRNA NCBI BLAST. The isolate's ambiguous clustering with both *Escherichia* and *Shigella* in Sourmash and phylogenetic analyses (Figures 2 and 3) highlights the limitations of relying solely on genomic approaches for distinguishing closely related species.

As several genomic approaches failed to yield a definitive and consistent taxonomic classification of the isolates, the biochemical and chromogenic differences among *E. coli*, *E. fergusonii*, and *Shigella* species on CCA were used to confirm the isolates' identity. The isolate's unique dark blue/purple appearance on CCA confirmed it as an *E. coli* strain, distinguishing it from *E. fergusonii* and *Shigella* species (Figure 6). This highlights the importance of

simple laboratory biochemical and/or cultural procedures for identifying and differentiating *E. coli* from *Shigella* species, as well as from *E. fergusonii.*

E. coli, a coliform bacterium, ferments lactose and appears dark blue/purple on CCA. *E. fergusonii*, on the other hand, is also a coliform but does not ferment lactose, appearing pink on CCA due to its inability to hydrolyse X-glucuronide in the medium. *E. coli* and *E. fergusonii* are closely related bacteria, both capable of causing illnesses in humans, such as bacteraemia, urinary tract infections, and diarrhoea (Lindsey et al., 2017). *Shigella* species are non-coliform bacteria, do not ferment lactose, and appear colourless on CCA as they do not hydrolyse either Salmon-GAL or X-glucuronide in the medium. *E. fergusonii* has long been recognized as a close relative of bacteria in the *Escherichia coli-Shigella* group (Farmer et al., 1985), and the genetic similarities between *E. coli, E. fergusonii,* and *Shigella* species may present challenges for clinical diagnosis and effective treatment if phenotypic approaches, such as biochemical and cultural properties, are neglected. Therefore, integrating traditional microbiological approaches, including culturing and biochemical characterization methods, alongside MALDI-TOF mass spectrometry, to support the genomic identification of clinically significant bacteria is essential.

CONCLUSION

This study describes the application of phenotypic and genomic approaches for the identification and antimicrobial characterization of an MDR bacterial strain, *E. coli* 266631E, isolated from the blood of a sepsis patient. Initially identified as *E. coli* through MALDI-TOF analysis, the isolate exhibited an alarming MDR profile after phenotypic antimicrobial screening (disk diffusion testing and broth microdilution MIC determination), prompting genomic studies to confirm its taxonomic identity and characterize its ARGs.

Genomic analysis revealed multiple ARGs encoding resistance to a broad range of clinically important antibiotics across different classes, including macrolides, fluoroquinolones, aminoglycosides, carbapenems, and cephalosporins. However, attempts to resolve the isolate's taxonomy by querying its 16S rRNA and genome sequences in recognised bacterial taxonomic databases, such as NCBI, Greengenes2, and SILVA, were inconclusive. The conserved nature of the 16S rRNA gene and the close phylogenetic relationship between *Shigella* and *Escherichia* hindered definitive species-level classification.

Although *E. coli* is extensively studied, its taxonomic classification remains challenging due to high genetic similarity with species like *Shigella* and *E. fergusonii*, particularly in conserved genomic regions—despite their phenotypic differences and variations in pathogenic traits. The isolate's ambiguous clustering with both *Escherichia* and *Shigella* in Sourmash and phylogenetic analyses highlights the limitations of relying solely on genomic methods for distinguishing closely related taxa. Ultimately, the species identity of *E. coli* 266631E was confirmed through a simple phenotypic assay using CCA, emphasizing the continued relevance of traditional microbiological techniques alongside modern genomic tools.

This study underscores the importance of integrating phenotypic methods with genomic analyses for accurate identification of clinically significant bacteria. The isolation of an MDR strain from a sepsis patient reinforces the urgent need for diagnostic strategies that capture both genetic and phenotypic resistance, enabling timely and targeted antimicrobial therapy in life-threatening infections.

SUPPLEMENTARY MATERIAL

The outputs of the MALDI-TOF analysis, AMR analysis using AMRFinderPlus, starAMR, and RGI, the query of bacterial 16S rRNA in the NCBI, Greengenes2, and SILVA databases, as well as Sourmash and pangenome analyses, are available in the supplementary material.

DATA AVAILABILITY

The bacterial 16S rRNA gene sequence has been deposited in the NCBI GenBank database under accession number PQ871642. The isolate's complete genome sequence has been submitted to the NCBI Genome database under BioProject accession PRJNA1220687 and BioSample accession SAMN46722626. The chromosome is available under GenBank accession number CP183489, while plasmids and other contigs are available under accession numbers CP183490–CP183498. The scripts used in the following bioinformatics analysis – Sourmash, Roary, and Prokka (for pangenome analysis) are available at: https://github.com/LucyDillon/MDR_isolate.

REFERENCES

1. Abram K, Udaondo Z, Bleker C, Wanchai V, Wassenaar TM, Robeson MS 2nd, Ussery DW. Mash-based analyses of *Escherichia coli* genomes reveal 14 distinct phylogroups. *Commun Biol* 2021; 4:117. https://doi.org/10.1038/s42003-020-01626-5
2. AL-Khikani F, Abdullah H, Karkaz H, abdulhussein, H. Evaluation the combination of amoxyclav with amikacin and ceftriaxone against *Escherichia coli* sepsis. *Microbes and Infectious Diseases*, 2023. https://doi.org/10.21608/mid.2023.206746.1516
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*, 1990; 215(3): 403–410. https://doi.org/10.1016/s0022-2836(05)80360-2
4. Bonten M, Johnson JR, van den Biggelaar AHJ, et al. Epidemiology of Escherichia coli bacteremia: a systematic literature review. Clin Infect Dis 2021; 72:1211–9. https://doi.org/10.1093/cid/ciaa210
5. Brown CT and Irber L. Sourmash: a library for MinHash sketching of DNA. *Journal of Open Source Software,* 2016; 1(5): 27. https://doi.org/10.21105/joss.00027
6. Buetti N, Marschall J, Atkinson A, Kronenberg A; Swiss Centre for Antibiotic Resistance (ANRESIS). National bloodstream infection surveillance in Switzerland 2008–2014: different patterns and trends for university and community hospitals. Infect Control Hosp Epidemiol 2016; 37:1060–7. https://doi.org/10.1017/ice.2016.137
7. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021; 38(12):5825-5829. https://doi.org/10.1093/molbev/msab293
8. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014; 58(7):3895-903. https://doi.org/10.1128/AAC.02412-14
9. Cheasty T., Rowe B. Antigenic relationships between the enteroinvasive *Escherichia coli* O antigens O28ac, O112ac, O124, O136, O143, O144, O152, and O164 and Shigella O antigens. *J. Clin. Microbiol.* 1983; 17:681–684. https://doi.org/10.1128/jcm.17.4.681-684.1983
10. Chukwunwejim CR, Eze PM, Ujam NT, Abonyi IC, Ejikeugwu CP, Abonyi DO, Esimone CO. Incidence of community acquired ESBL-producing bacteria among asymptomatic University students in Anambra State, Nigeria. *European Journal of Biological Research,* 2018; 8(3):138-147. http://dx.doi.org/10.5281/zenodo.1314719
11. Clinical and Laboratory Standards Institute (CLSI). Principles and procedures for blood cultures; approved guideline. 2nd ed. CLSI document M47-E2. Wayne (PA): CLSI; 2022.
12. Doua J, Geurtsen J, Rodriguez-Baño J, Cornely OA, Go O, Gomila-Grange A, Kirby A, Hermans P, Gori A, Zuccaro V, Gravenstein S, Bonten M, Poolman J, Sarnecki M. Epidemiology, Clinical Features, and Antimicrobial Resistance of Invasive *Escherichia Coli* Disease in Patients Admitted in Tertiary Care Hospitals. *Open Forum Infectious Diseases*, 2023, 10(2). https://doi.org/10.1093/ofid/ofad026
13. Dugar S, Choudhary C, Duggal A. Sepsis and septic shock: guideline-based management. Cleve Clin J Med 2020; 87:53–64. https://doi.org/10.3949/ccjm.87a.18143
14. Edgar R. Taxonomy annotation and guide tree errors in 16S rRNA databases. PeerJ. 2018, 6:e5030. https://doi.org/10.7717/peerj.5030
15. Elliott E, Brink AJ, van Greune J, Els Z, Woodford N, Turton J, Warner M, Livermore DM. In Vivo Development of Ertapenem Resistance in a Patient with Pneumonia Caused by *Klebsiella pneumoniae* with an Extended-Spectrum β-Lactamase, *Clinical Infectious Diseases*, 2006; 42(11): e95–e98, https://doi.org/10.1086/503264
16. EUCAST. Routine and extended internal quality control for MIC determination and disk diffusion as recommended by EUCAST. The European Committee on Antimicrobial Susceptibility Testing. Version 13.0, 2023. http://www.eucast.org
17. Farmer JJ 3rd, Fanning GR, Davis BR, O'Hara CM, Riddle C, Hickman-Brenner FW, Asbury MA, Lowery VA 3rd, Brenner DJ. Escherichia fergusonii and Enterobacter taylorae, two new species of Enterobacteriaceae isolated from clinical specimens. J Clin Microbiol. 1985; 21(1):77-81. https://doi.org/10.1128/jcm.21.1.77-81.1985

18. Gashaw M, Ali S, Berhane M, Tesfaw G, Eshetu B, Workneh N, Seeholzer T, Froeschl G, Kroidl A, Wieser, A, Gudina EK. Neonatal Sepsis Due to Multidrug-resistant Bacteria at a Tertiary Teaching Hospital in Ethiopia. The Pediatric Infectious Disease Journal, 2024; 43(7): 687-693. https://doi.org/10.1097/INF.0000000000004364
19. Geurtsen J, de Been M, Weerdenburg E, Zomer A, McNally A, Poolman J. Genomics and pathotypes of the many faces of Escherichia coli. FEMS Microbiol Rev. 2022; 46(6): fuac031. https://doi.org/10.1093/femsre/fuac031
20. Grevskott DH, Salvà-Serra F, Moore ERB, Marathe NP. Escherichia coli novel sequence type 11873 harbours a new CTX-M-15-carrying multidrug resistance type 1/2 hybrid IncC plasmid. J Glob Antimicrob Resist. 2024; 39:109-113. https://doi.org/10.1016/j.jgar.2024.08.008
21. Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086
22. Hsieh Y-P, Hung Y-M, Tsai M-H, Lai L-C and Chuang EY (2022) 16S-ITGDB: An Integrated Database for Improving Species Classification of Prokaryotic 16S Ribosomal RNA Sequences. *Front. Bioinform.* 2:905489. https://doi.org/10.3389/fbinf.2022.905489
23. Janda JM, Abbott SL. 2007. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J Clin Microbiol 45: https://doi.org/10.1128/jcm.01228-07
24. Jenkins C, Ling CL, Ciesielczuk HL, Lockwood J, Hopkins S, McHugh TD, Gillespie SH, Kibbler CC. Detection and identification of bacteria in clinical samples by 16S rRNA gene sequencing: comparison of two different approaches in clinical practice. J Med Microbiol. 2012; 61(Pt 4):483-488. https://doi.org/10.1099/jmm.0.030387-0
25. Jo JH, Kennedy EA, Kong HH. Research Techniques Made Simple: Bacterial 16S Ribosomal RNA Gene Sequencing in Cutaneous Research. J Invest Dermatol. 2016; 136(3): e23-e27. https://doi.org/10.1016/j.jid.2016.01.005
26. Khot PD, Fisher MA 2013. Novel Approach for Differentiating Shigella Species and Escherichia coli by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry. J Clin Microbiol 51(11). https://doi.org/10.1128/jcm.01526-13
27. Kumar NR, Balraj TA, Kempegowda SN, Prashant A. Multidrug-Resistant Sepsis: A Critical Healthcare Challenge. *Antibiotics*. 2024; 13(1):46. https://doi.org/10.3390/antibiotics13010046
28. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32(1): 268-74 https://doi.org/10.1093/molbev/msu300
29. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect. Immun., 2004; 72(9): 5080-5088. https://doi.org/10.1128/iai.72.9.5080-5088.2004
30. Lan R, Reeves PR. When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol. 2001; 9(9): 419-24. https://doi.org/10.1016/s0966-842x(01)02133-3
31. Lan R, Lumb B, Ryan D, Reeves PR. Molecular evolution of large virulence plasmid in Shigella clones and enteroinvasive *Escherichia coli*. *Infect. Immun.* 2001; 69:6303–6309. https://doi.org/10.1128/iai.69.10.6303-6309.2001
32. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024; 52(W1): W78-W82. https://doi.org/10.1093/nar/gkae268
33. Lindsey RL, Garcia-Toledo L, Fasulo D, Gladney LM, Strockbine N. Multiplex polymerase chain reaction for identification of Escherichia coli, Escherichia albertii and Escherichia fergusonii. J Microbiol Methods. 2017; 140: 1-4. https://doi.org/10.1016/j.mimet.2017.06.005
34. Liu W, Tang J-W, Mou J-Y, Lyu J-W, Di Y-W, Liao Y-L, Luo Y-F, Li Z-K, Wu X and Wang L (2023) Rapid discrimination of *Shigella* spp. and *Escherichia coli via* label-free surface enhanced Raman spectroscopy coupled with machine learning algorithms. *Front. Microbiol.* 14:1101357. https://doi.org/10.3389/fmicb.2023.1101357

35. Livermore DM, Hawkey PM. CTX-M: changing the face of ESBLs in the UK, *Journal of Antimicrobial Chemotherapy*, 2005; 56(3): 451–454. https://doi.org/10.1093/jac/dki239
36. Madrazo M, López-Cruz I, Piles L, Viñola S, Alberola J, Eiros JM, Artero A. Risk Factors and the Impact of Multidrug-Resistant Bacteria on Community-Acquired Urinary Sepsis. *Microorganisms*. 2023; 11(5):1278. https://doi.org/10.3390/microorganisms11051278
37. Maldonado N, López-Hernández I, López-Cortés LE, Pérez-Crespo PM, Retamar-Gentil P, García-Montaner A, et al. Association of microbiological factors with mortality in *Escherichia coli* bacteraemia presenting with sepsis/septic shock: a prospective cohort study. Clinical Microbiology and Infection, 2024, 30(8): 1035-1041. https://doi.org/10.1016/j.cmi.2024.04.001
38. Markwart, R.; Saito, H.; Harder, T.; Tomczyk, S.; Cassini, A.; Fleischmann-Struzek, C.; Reichert, F.; Eckmanns, T.; Allegranzi, B. Epidemiology and Burden of Sepsis Acquired in Hospitals and Intensive Care Units: A Systematic Review and Meta-Analysis. *Intensive Care Med.* 2020, *46*, 1536–1551. https://doi.org/10.1007/s00134-020-06106-2
39. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, Morton JT, et al. Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol 42, 715–718 (2024). 10.1038/s41587-023-01845-1
40. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266
41. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 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
42. Oyama LB, Olleik H, Teixeira ACN, et al. In silico identification of two peptides with antibacterial activity against multidrug-resistant Staphylococcus aureus. *NPJ Biofilms Microbiomes*. 2022; 8(1):58. https://doi.org/10.1038/s41522-022-00320-0
43. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 2015; 31(22): 3691–3693. https://doi.org/10.1093/bioinformatics/btv421
44. Peng Z, Wang X, Huang J, Li B. Chapter 53 - Pathogenic *Escherichia coli.* In: Molecular Medical Microbiology (Third Edition) (Editors: Tang Y-W, Hindiyeh MY, Liu D, Sails A, Spearman P, Zhang J-R). Academic Press, 2024, Pages: 1065-1096. https://doi.org/10.1016/B978-0-12-818619-0.00069-1
45. Pérez-Crespo PMM, Lanz-García JF, Bravo-Ferrer J, et al. Revisiting the epidemiology of bloodstream infections and healthcare-associated episodes: results from a multicentre prospective cohort in Spain (PRO-BAC study). Int J Antimicrob Agents 2021; 58:106352. https://doi.org/10.1016/j.ijantimicag.2021.106352
46. Pettengill EA, Pettengill JB and Binet R. Corrigendum: Phylogenetic Analyses of *Shigella* and Enteroinvasive *Escherichia coli* for the Identification of Molecular Epidemiological Markers: Whole-Genome Comparative Analysis Does Not Support Distinct Genera Designation. *Front. Microbiol*., 2018; 8:2598. https://doi.org/10.3389/fmicb.2017.02598
47. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009; 26(7):1641-50. https://doi.org/10.1093/molbev/msp077
48. Priyanka P, Meena PR, Raj D, Rana A, Dhanokar A, Duggirala KS, Singh AP. Urinary tract infection and sepsis causing potential of multidrug-resistant Extraintestinal pathogenic E. coli isolated from plant-origin foods. International Journal of Food Microbiology, 2023, 386: 110048. https://doi.org/10.1016/j.ijfoodmicro.2022.110048
49. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics, 2012; 28: 1823-1829. https://doi.org/10.1093/bioinformatics/bts252
50. Pupo GM, Lan R, Reeves PR. Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. Proc. Natl. Acad. Sci. U. S. A., 2000; 97(19):10567-10572. https://doi.org/10.1073/pnas.180094797
51. Rangama S, Lidbury IDEA, Holden JM, Borsetto C, Murphy ARJ, Hawkey PM, Wellington EMH. Mechanisms Involved in the Active Secretion of CTX-M-15 β-Lactamase by Pathogenic Escherichia coli

ST131. Antimicrob Agents Chemother., 2021. 65: 10.1128/aac.00663-21. https://doi.org/10.1128/aac.00663-21

52. Rantakokko-Jalava K, Nikkari S, Jalava J, Eerola E, Skurnik M, Meurman O, Ruuskanen O, Alanen A, Kotilainen E, Toivanen P, Kotilainen P. Direct amplification of rRNA genes in diagnosis of bacterial infections. J Clin Microbiol. 2000, 38(1):32-9. https://doi.org/10.1128/JCM.38.1.32-39.2000
53. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom. 2018; 4(8): e000206. https://doi.org/10.1099/mgen.0.000206
54. Rosenberg S, Bonten M, Haazen W, Spiessens B, Abbanat D, Go O, Wagenlehner F, Shore N, Hagiwara Y, Ibarra de Palacios P, Geurtsen J, Hermans P, Poolman J. Epidemiology and O-serotypes of extraintestinal pathogenic Escherichia coli disease in patients undergoing transrectal ultrasound prostate biopsy: a prospective multicenter study. J Urol 2021; 205:826–32. https://doi.org/10.1097/ju.0000000000001425
55. Rossolini GM, D'Andrea MM, Mugnaioli C. The spread of CTX-M-type extended-spectrum beta-lactamases. Clin Microbiol Infect. Clin Microbiol Infect. 2008; 14: 33 - 41. https://doi.org/10.1111/j.1469-0691.2007.01867.x
56. Sanford RA, Lloyd KG, Konstantinidis KT, Löffler FE. Microbial Taxonomy Run Amok. Trends Microbiol. 2021; 29(5):394-404. https://doi.org/10.1016/j.tim.2020.12.010
57. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 2014; 30(14): 2068–2069. https://doi.org/10.1093/bioinformatics/btu153
58. Shao Q, Chen D, Chen S, Ru X, Ye Q. *Escherichia coli* Infection Sepsis: An Analysis of Specifically Expressed Genes and Clinical Indicators. *Diagnostics*. 2023; 13(23):3542. https://doi.org/10.3390/diagnostics13233542
59. Sigma-Aldrich. 81938 Coliform ChromoSelect Agar. https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/392/035/81938dat.pdf?srsltid=AfmBOoo2hlPknMoo53IfShsZESU5w_dRvUf7sg61QGo5Q7tx-sPACqUX Date accessed: 01.10.2024
60. Song K, Guo C, Zeng Z. *et al.* Factors associated with in-hospital mortality in adult sepsis with *Escherichia coli* infection. *BMC Infect Dis* 2022; 22:197. https://doi.org/10.1186/s12879-022-07201-z
61. Stoll BJ, Puopolo KM, Hansen NI, Sánchez PJ, Bell EF, Carlo WA, Cotten CM, D'Angio CT, Kazzi SNJ, Poindexter BB, et al. Early-Onset Neonatal Sepsis 2015 to 2017, the Rise of Escherichia coli, and the Need for Novel Prevention Strategies. JAMA Pediatr. 2020; 174(7): e200593. https://doi.org/10.1001/jamapediatrics.2020.0593
62. The HC, Bodhidatta L, Pham DT, Mason CJ, Ha Thanh T, Voong Vinh P, Turner P, Hem S, Dance DAB, Newton PN, Phetsouvanh R, Davong V, Thwaites GE, Thomson NR, Baker S, Rabaa MA. Evolutionary histories and antimicrobial resistance in Shigella flexneri and Shigella sonnei in Southeast Asia. Commun Biol. 2021, 19;4(1):353. https://doi.org/10.1038/s42003-021-01905-9
63. The H, Thanh D, Holt K. *et al.* The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol.* 2016, 14, 235–250. https://doi.org/10.1038/nrmicro.2016.10
64. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017; 13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595
65. Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, Xue Y, Zhu Y, Xu X, Sun L, Chen S, Nie H, Peng J, Xu J, Wang Y, Yuan Z, Wen Y, Yao Z, Shen Y, Qiang B, Hou Y, Yu J, Jin Q. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. 2005; 33(19):6445-58. https://doi.org/10.1093/nar/gki954
66. Yu K, Huang Z, Xiao Y, Bai X, Gao H, Wang D. Epidemiology and molecular characterization of CTX-M-type ESBLs producing Escherichia coli isolated from clinical settings. J Glob Antimicrob Resist. 2024; 36:181-187. https://doi.org/10.1016/j.jgar.2023.11.013