GMR
The Original

COMPARATIVE GENOMIC ANALYSIS OF MULTIDRUG-RESISTANT BACTERIAL ISOLATES FROM CLINICAL AND ENVIRONMENTAL SOURCES

Dr. Daniel Ningthoujam[1*], Dr Ashangbam Vivekananda Singh[2], Dr. Vinod Kumar[3], Dr. Shuchi Gupta[4], S. Karthik Sundaram[5], Md Tausif Raza[6]

[1*]Assistant Professor, Department of Microbiology, Shija Academy of Health Sciences, Langol, Imphal West, Manipur, Email ID: ningthoujam12@gmail.com, ORCID ID: 0000-0003-1462-292X
[2]Associate Professor, Department of Microbiology, Mahabodhi Medical College and Hospital, Gaya, Bihar, Email Id: avivekanandasingh@gmail.com
[3]Laboratory Technician, Postgraduate Department of Microbiology (VRDL-ICMR), Govt. Medical College, Jammu, E-mail ID: vkvkumar851@gmail.com, ORCID ID: 0009-0009-3165-6583
[4] Associate Professor, Microbiology, College of Nursing and Paramedical, JNCTPU, Bhopal (M.P.), E-mail ID: shuchigupta.micro@gmail.com, ORCID ID: 0009-0006-1683-4159
[5]Associate Professor, Department of Microbiology, Dr. N.G.P. Arts and Science College, Coimbatore, Email - karthiksundaram@drngpasc.ac.in, ORCID ID: 0000-0002-5964-2740
[6] a Microbiologist, RT PCR Lab, Sadar Hospital, Siwan, Bihar 841226,
b Research Scholar, Department of Microbiology, Smt S S Nootan Patel Science and Commerce College, Sankalchand Patel University, Visnagar, Mahesana, Gujarat, Email ID: tausifmicro@gmail.com, ORCID ID: 0009-0007-6583-2421

ABSTRACT:
Multidrug-resistant *Escherichia coli* is found in clinical and environmental settings, and there is a need to understand the variation in the range of resistances, genomic lineage or mobile genetic elements between these settings. This study compared six multidrug-resistant *E. coli* genome assemblies comprising three clinical and three environmental isolates. Comparative whole-genome bioinformatics included genome characterisation, multilocus sequence typing, acquired antimicrobial resistance gene (ARG) profiling, resistance-class analysis, plasmid replicon typing, Mash-based genomic relatedness, and ARG Jaccard similarity. Environmental genomes carried 11–16 unique ARGs compared with 8–10 in clinical genomes and collectively contained 25 distinct ARGs versus 21 among clinical isolates. Eight ARGs were shared between sources, with an overall source-level Jaccard similarity of 0.211. ST69 occurred in both clinical and environmental isolates, while the closest genomic pair also comprised isolates from different sources. Mash distance showed a negative but non-significant association with resistome similarity (Spearman $\rho = -0.500$, $P = 0.0577$). Plasmid replicon richness ranged from one to seven per genome, with IncFIB(AP001918) being the most widely distributed replicon, while replicon richness was not associated with ARG burden ($\rho = -0.058$, $P = 0.9131$). These results show both unique and overlapping resistomes and genomic relatedness across multiple sources, providing a basis for genomic surveillance of antimicrobial resistance integrated across clinical and environmental compartments.

KEYWORDS: antimicrobial resistance, *Escherichia coli*, comparative genomics, resistome, plasmid replicons

1. INTRODUCTION

A major challenge to effective management of infectious diseases is antimicrobial resistance (AMR), which diminishes the therapeutic potential of existing antimicrobial agents and increases the risk of failure. The evolution of resistance to an individual antibiotic is not the only issue, since the bacteria may be able to acquire resistance determinants to more than one class of antibiotics. Multidrug resistance makes treatment more complicated and limits the options (Tang *et al*., 2023). The rise of resistant bacteria is a result of the continual exposure of microbes to antimicrobial agents, selection pressure, adaptation, and spread of resistance genes from one microbe to another. These processes highlight that AMR is a biological and ecological issue and not just antimicrobial prescribing (Amann *et al*., 2019). This has long-term impacts, such as longer infections, Health spending, strains on health systems, and increased risk of normally treatable bacterial infections becoming difficult to treat (Ahmed *et al*., 2024). Amongst the multidrug-resistant Gram-negative bacteria are those with very adaptive acquired resistance mechanisms and intrinsic characteristics. Clinically important Gram-negative pathogens may harbor resistance mechanisms against multiple classes of antibiotics, including β-lactams, aminoglycosides, fluoroquinolones, polymyxins and other classes, and therefore be resistant to a wide range of antibiotics with narrow therapeutic indices (Macesic *et al*., 2025). They also have the ability to develop resistance through plasmids and other mobile genetic elements, which aids in the development of multidrug-resistant lineages. These organisms are also clinically relevant due to their persistence, ability to colonize hosts and spread of resistance genes among bacterial populations (Exner *et al*., 2017).

Escherichia coli is an informative model organism to study these processes as it is found in both human-associated and environmental settings and contains lineages that have very different resistance patterns. Resistance can be developed by changing the target, by inactivating the antimicrobial, by changing the permeability structure of the target, by inhibiting the target's ability to pump the antimicrobial out and by getting new genes that carry resistance (Moo *et al*., 2020). Multi-drug-resistant and extended-spectrum β-lactamase-producing *E. coli* (ESBLs) are clinical isolates that are

important in human infection, as shown by clinical investigations, and their acquired resistance determinants can continue to accumulate within these isolates (Ibrahim *et al*., 2023). The genomic versatility of *E. coli* renders it useful for studying the presence of resistance traits in environmental reservoirs and to determine if they are the same as those seen in clinical populations. Increasingly, aquatic ecosystems are seen as a part of the broader AMR picture. Multidrug-resistant *E. coli* have been isolated from drinking-water sources, indicating that there is no restriction of resistant populations to healthcare facilities and that they can be found in water systems that are directly relevant to human exposure (Odonkor & Addo, 2018). Wastewater and nearby human activities can introduce bacteria and antimicrobial residues, resistance genes and other selective agents into the water of rivers. The presence of resistance determinants in the wastewater-river systems suggests opportunities for the coexistence of differently sourced resistant bacteria, due to resistance determinants being able to persist and spread through connected water bodies (Read *et al*., 2024). Water can therefore be a reservoir and recipient of antibiotic-resistant bacteria, which can help resistance to be maintained and redistributed beyond their original host (Meradji *et al*., 2025). AMR is interconnected to humans and the environment, which can be interpreted as One Health. Resistant organisms and their genetic determinants can be shared among humans, animals, and environments, which can result in a surveillance system that is limited to a single compartment and may not be representative of the resistance ecology (Al-Khalaifah *et al*., 2025). Comparative genomic analysis can be used to look at this connection at a finer scale. Whole-genome sequencing (WGS) can also be used to help with lineage assignment, characterization of resistance genes, analysis of mobile genetic elements, and measuring genomic relatedness and is becoming increasingly useful for AMR surveillance and for conducting an epidemiological investigation (Matsumura *et al*., 2025).

One of the challenges in the current studies is the fact that both clinical and environmental multidrug-resistant *E. coli* are frequently described in their original environment and not in a unified genomic context. Resistance phenotype does not necessarily reflect close genomic relatedness since antimicrobial resistance gene(s) can be acquired horizontally by genetically distinct isolates. On the other hand, because of the gain or loss of the mobile resistance elements, the acquired resistomes of closely related genomes can vary. Therefore, together with sequence type, whole-genome relatedness, distribution of resistance classes and plasmid replicon profiles, evaluation of the antimicrobial resistance gene composition can be useful in delineating similarities between different sequence types that could reflect bacterial genomic background and those that may reflect mobile resistance determinants.

The overall objective of the present study was to comparatively characterize the multidrug-resistant *E. coli* genomes from clinical and environmental origins. Specific objectives were to compare the gene repertoires for antimicrobial resistance acquired from the two sources, to determine sequence types and whole-genome relatedness and to examine the correlation between the resistome similarity and the plasmid replicon diversity and the relationship with the burden of acquired resistance genes. This holistic approach was developed to enable focused genomic evaluation of common and source-specific resistance traits on the clinical-environmental interface.

2. METHODOLOGY

2.1 Research Design

The multidrug-resistant Escherichia coli (*MDR E. coli*) from various clinical and environmental sources were characterized using a comparative whole-genome bioinformatics approach. Three clinical and three environmental isolates were used for analysis and were considered as 6 genome assemblies. This workflow involved genome assembly characterization, multilocus sequence typing (*MLST*), detection of acquired antimicrobial resistance genes (ARG), resistome classification, plasmid replicon typing, whole-genome relatedness, and resistome similarity. The outputs from the genome level were kept as outputs with the corresponding accession numbers to facilitate consistent integration across the various analyses.

2.2 Data Source

Genome assemblies were retrieved from two public BioProjects containing multidrug-resistant *E. coli* genomic data. Multiple genomes of *E. coli* were downloaded from the BioProject PRJNA605141, which comprises genome sequencing and assembly data of *E. coli* isolates with multiple phenotypic antimicrobial resistance and is linked to the isolates described by Neumann *et al*. (2020). The environmental genomes were downloaded from BioProject PRJNA756839, which includes multidrug-resistant *E. coli* from river environments related to the study of Hara *et al*. (2018). Three assemblies were chosen from each source for downstream comparative genomics and the accession remained constant throughout the process.

2.3 Genome Characterization and Multilocus Sequence Typing

For the genome assemblies, the total genome size, GC content, number of contigs, largest contig length, N50, and L50 were used as genome assembly-level descriptors. Multilocus sequence typing (MLST) was performed using MLST v2.35.0, applying the Achtman *E. coli* MLST scheme to assign sequence types (STs) from the assembled genomes. The seven housekeeping loci adk, fumC, gyrB, icd, mdh, purA and recA were used to determine their allelic profiles in order to sequence the types. While genomes with an unresolved allele were kept as incomplete to prevent unsupported assignment of a sequence type, only those with complete allele assignments were assigned a definitive sequence type. Both of these analyses gave the genomic and lineage characteristics that were used as the basis for comparative analysis of clinical and environmental isolates.

2.4 Antimicrobial Resistance Gene Profiling

The acquired ARGs were detected by using ResFinder screening of genome assemblies. ARG names were standardized to unique ARG names in each genome to avoid double-detections and overestimate the resistome burden. The resulting presence-absence matrix was then put into use for calculating the number of unique ARGs per isolate and to determine ARGs shared between clinical and environmental genomes, as well as ARGs found in only one genome source. ARGs were also classified by antimicrobial resistance classes to be able to compare the antimicrobial resistance gene-level diversity and representation of the two source groups.

The mean number of ARGs and ARG median burden, range of ARGs, total number of ARGs, and mean number of the ARGs belonging to different resistance classes per genome were used to summarise the source-level resistome diversity. The overlap in resistome composition was quantified by calculating the Jaccard similarity based on the presence–absence of binary data for the ARG. These measures included each unique ARG only once per genome; duplicate detections were not treated as independent observations.

2.5 Plasmid Replicon Analysis

Plasmid-associated replicon sequences were searched with PlasmidFinder software. Duplicate calls within each genome were removed, and replicon names were standardized before comparative analysis. The number of unique replicon types found per genome was defined as replicon richness. Then, presence–absence profiles were compared between clinical and environmental isolates and analyzed together with the MLST assignments to look for patterns in plasmid backgrounds over the distribution of genomic lineages. The richness of replicons detected in each of the genomes was also included and compared to the ARG burden to determine if the more diverse the number of detectable replicons, the more extensive the acquired resistome.

2.6 Whole-Genome Relatedness and Resistome Similarity

The relatedness of the genomes was estimated with Mash v2.3. Each assembly was used to create genome sketches, and Mash distances between all 15 unique genome pairs were determined. These were clustered in a hierarchical approach to show the relationship between clinical and environmental isolates using the distance matrix produced. Subsequently, the pairwise comparisons were divided into two groups: within source and between source.

A pairwise integration of mash distance and ARG similarity was performed to test whether there was a connection between these two variables for more similar acquired resistance genes in more similar genomes. This method allowed for the assessment of genomic relatedness and resistome concordance without taking isolation source into account.

2.7 Statistical and Comparative Analysis

Genome characteristics, ARG burden, resistance-class diversity and plasmid replicon richness were calculated as descriptive statistics. A source-level summary was used to compare clinical and environmental genomes, as there were three isolates in each group. The correlation between the ARG Jaccard similarity and pairwise Mash distance, as well as between the plasmid replicon richness and acquired ARG burden, was measured using Spearman's rank correlation. A P value < 0.05 was considered statistically significant. Because there are only a few genomes, correlation analyses were considered exploratory only, and not as estimates at the population level.

3. RESULTS

3.1 Genome characteristics and sequence typing

There were three clinical and three environmental strains of Escherichia coli. Genome sizes ranged from 4.88 to 5.46 Mb, with GC contents of 50.40–50.81%. The size of the contigs ranged from 3 to 204 contigs per assembly, and N50 values ranged from 100,204 bp to 5,109,767 bp. The genome characteristics and MLST assignments are listed in Table 1. Five isolates yielded complete Achtman sequence types; ST131, ST69, and ST155 were from clinical isolates, and ST69 and ST4162 were from environmental isolates. The allelic profile was incomplete for GCA_022213095.1, and its sequence type remained unresolved. ST69 was identified in both clinical and environmental isolates.

Table 1. Genome assembly characteristics and multilocus sequence types of clinical and environmental *E. coli* isolates

Accession	Source	Genome size (bp)	GC (%)	Contigs	Largest contig (bp)	N50 (bp)	L50	Sequence type	MLST status
GCA_000285655.3	Clinical	5,249,449	50.81	3	5,109,767	5,109,767	1	131	Complete
GCA_010806885.1	Clinical	5,314,319	50.59	192	473,257	145,421	12	69	Complete
GCA_010806905.1	Clinical	4,921,290	50.60	154	317,850	119,280	13	155	Complete

GCA_022213085.1	Environmental	5,460,747	50.46	204	515,674	100,204	15	69	Complete
GCA_022213095.1	Environmental	5,198,255	50.40	149	501,939	109,874	13	Unresolved	Incomplete
GCA_022213145.1	Environmental	4,882,886	50.59	114	314,499	139,964	12	4162	Complete

3.2 Acquired antimicrobial resistance gene profiles

All six genomes contained acquired ARGs, with 8-10 ARGs found in the clinical and 11-16 in the environmental genomes. Table 2 displays ARG profiles per isolate, with GCA_022213095.1 having the most ARG genes (n = 16). The genomes contained 6-8 resistance classes each. Individual ARGs were distributed as follows: tet(A) was found in all of the isolates, and blaTEM-1B was found in five. Multiple determinants with source-restricted distribution were found in the analyzed genome set, such as tet(X4), fosA3/fosA4, qnrS1 (environmental genome) and mcr-1.1/mcr-1.27 (clinical genome). Dark cells denote ARG detection, light cells denote absence, and genome labels are coloured based on source of isolation.

Table 2. Acquired antimicrobial resistance gene profiles of clinical and environmental *E. coli* genomes

Accession	Source	Unique ARGs	Resistance classes	Detected ARGs
GCA_000285655.3	Clinical	10	7	aac(6')-Ib-cr, aadA5, blaCMY-23, blaCTX-M-15, blaOXA-1, blaTEM-1B, dfrA17, mph(A), sul1, tet(A)
GCA_010806885.1	Clinical	8	8	ant(3")-Ia, blaTEM-1B, catA1, dfrA1, mcr-1.27, mph(B), sul1, tet(A)
GCA_010806905.1	Clinical	9	7	ant(3")-Ia, aph(3")-Ib, aph(6)-Id, blaTEM-1B, dfrA14, lnu(F), mcr-1.1, sul2, tet(A)
GCA_022213085.1	Environmental	11	6	aac(3)-IIa, ant(3")-Ia, aph(3")-Ib, aph(3')-Ia, aph(6)-Id, blaTEM-1B, dfrA14, FloR, sul2, tet(A), tet(X4)
GCA_022213095.1	Environmental	16	7	aac(3)-IVa, aadA2, ant(3")-Ia, aph(3')-Ia, aph(4)-Ia, blaTEM-176, CmlA7, dfrA12, FloR, fosA4, qnrS1, sul2, sul3, tet(A), tet(M), tet(X4)
GCA_022213145.1	Environmental	12	7	aac(3)-IVa, ant(3")-Ia, aph(4)-Ia, blaCMY-2, blaCTX-M-65, blaTEM-1B, dfrA1, FloR, fosA3, qnrS1, tet(A), tet(X4)

Figure 1. Distribution of acquired antimicrobial resistance genes across clinical and environmental Escherichia coli genomes.

3.3 Clinical and environmental resistome comparison

The mean ARG burden was higher in the environmental genomes compared to clinical genomes (13.0 ARGs/genome versus 9.0 ARGs/genome). The resistome comparison between the sources is presented in Table 3. The number of ARGs found in environmental isolates was 25, while 21 ARGs were found in clinical isolates. There were eight ARGs shared by the sources, 17 specific to environmental isolates, and 13 specific to clinical isolates. The overall ARG Jaccard similarity of source groups was 0.211. Figure 2A illustrates the ARG burden by source of isolation, where clinical genomes carried 8-10 ARGs and environmental genomes 11-16 ARGs. The prevalence of resistance is shown in Figure 2B, with aminoglycoside resistance, tetracycline resistance and trimethoprim resistance included in all genomes. Fosfomycin resistance was found in environmental genomes, while macrolide resistance, polymyxin resistance and lincosamide resistance were found in clinical genomes.

Table 3. Comparison of acquired antimicrobial resistance gene diversity between clinical and environmental *E. coli* genomes

Source	Genomes (n)	Mean ARGs/genome	Median ARGs/genome	ARG range	Distinct ARGs	Mean resistance classes/genome	Shared ARGs	Source-specific ARGs	Overall ARG Jaccard similarity
Clinical	3	9.0	9.0	8–10	21	7.33	8	13	0.211
Environmental	3	13.0	12.0	11–16	25	6.67	8	17	0.211

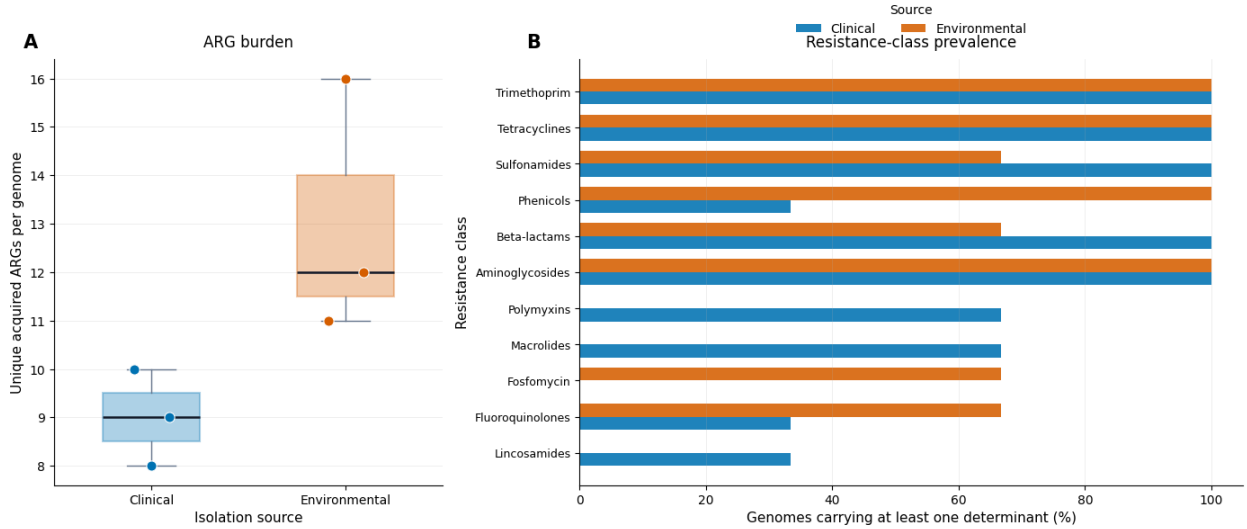

Figure 2. Comparative resistome burden and resistance-class distribution. (A) Unique acquired ARG burden in clinical and environmental *E. coli* genomes. (B) Prevalence of resistance classes according to isolation source

3.4 Whole-genome relatedness and resistome similarity

Within and between isolation sources, there was a tendency for clustering that was observed with genome-wide relatedness. The Mash-based genomic relationships generated in Figure 3A are closest between clinical GCA_010806905.1 and environmental GCA_022213145.1 (Mash distance = 0.004360). The clinical and environmental ST69 isolates, GCA_010806885.1 and GCA_022213085.1, were also relatively closely related (0.007232). Figure 3B shows the relationship between Mash distance and ARG Jaccard similarity. In 15 pairwise comparisons, resistome similarity was negatively correlated with Mash distance, but not statistically significant (Spearman $\rho = -0.500$, $P = 0.0577$). The means of ARG similarities were 0.2381 for within-source comparisons and 0.1795 for between-source comparisons.

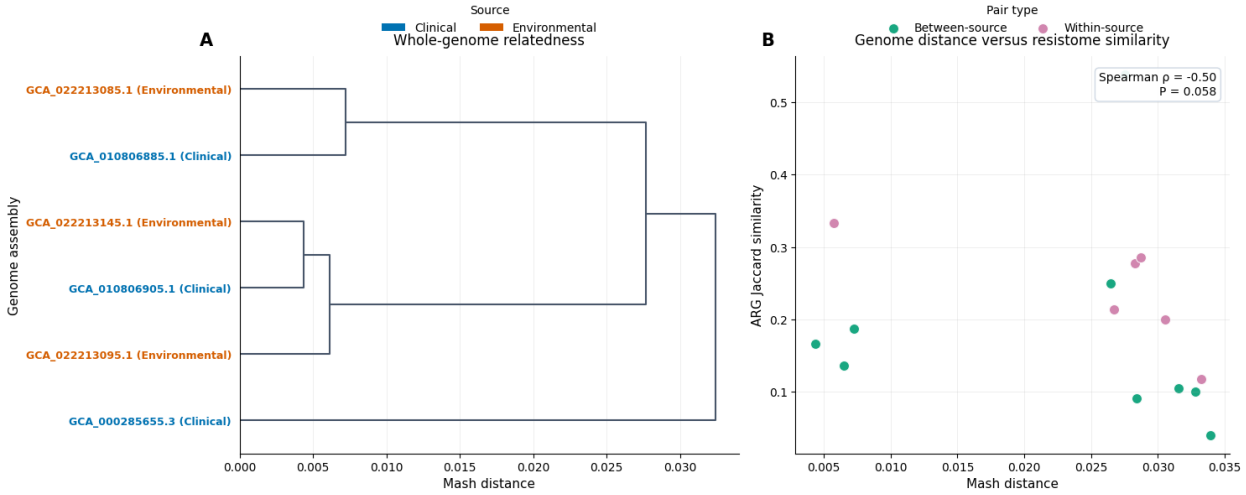

Figure 3. Whole-genome relatedness and resistome similarity. (A) Hierarchical clustering of clinical and environmental *E. coli* genomes based on Mash distances. (B) Relationship between whole-genome Mash distance and ARG Jaccard similarity for within-source and between-source genome pairs.

3.5 Plasmid replicon diversity and ARG burden

The number of plasmid replicons per genome ranged from 1 to 7 per genome. The plasmid replicon profiles are shown in Table 4, GCA_022213085.1 had the highest plasmid replicon richness (n = 7). The most widely distributed replicon was IncFIB(AP001918), which was found in five genomes. The plasmid replicon distribution is presented with respect to the MLST (Figure 4A), and the different replicon distributions are highlighted between the two ST69 genomes. Figure 4B shows the relationship between plasmid replicon richness and ARG burden. No statistically significant correlation was observed between replicon richness and ARG burden (Spearman $\rho = -0.058$, $P = 0.9131$).

Table 4. Plasmid replicon profiles of clinical and environmental *E. coli* genomes

Accession	Source	Unique replicons	Detected plasmid replicons
GCA_000285655.3	Clinical	2	IncFIA, IncFII
GCA_010806885.1	Clinical	6	Col(BS512), Col(MG828), Col(pHAD28), Col156, IncFIB(AP001918), IncX4
GCA_010806905.1	Clinical	5	Col(MG828), IncFIB(AP001918), IncFIC(FII),

			IncI1_1_Alpha, IncX4
GCA_022213085.1	Environmental	7	Col(MG828), Col(pHAD28), Col156, IncFIA, IncFIB(AP001918), IncN, IncX1
GCA_022213095.1	Environmental	6	IncFIA(HI1)_1_HI1, IncFIB(AP001918), IncFII, IncHI1A, IncHI1B(R27)_1_R27, IncX1
GCA_022213145.1	Environmental	1	IncFIB(AP001918)

Figure 4. Plasmid replicon diversity, MLST, and antimicrobial resistance burden. (A) Distribution of plasmid replicons across genomes with corresponding Achtman sequence types. (B) Relationship between unique plasmid replicon count and acquired ARG burden according to isolation source.

4. DISCUSSION

The genomic profile of the two groups (clinical and environmental) of *E. coli* showed there were large differences in the resistomes acquired by the two groups, although there was evidence of genomic relatedness between *E. coli* from all sources and across all source boundaries. The environmental isolates had a higher ARG burden with 11–16 ARG per genome, whereas among clinical isolates, there were 8–10 ARG per genome and a larger number of distinct ARGs in the environmental isolates. The low similarity of ARG structure indices, such as the Jaccard similarity, between the two groups also suggested that only a small fraction of the ARG structure of the two groups was shared. Such a pattern may indicate that the resistance profiles of clinical and environmental isolates of related genomic backgrounds are greatly divergent. Genes like tet(X4), qnrS1 and fosA3/fosA4 were found only in environmental isolates, while mcr-1.1 and mcr-1.27 were found only in clinical isolates, which provided additional evidence for the source-associated resistome structure. There was, however, much overlap between resistance-class profiles, especially for the aminoglycosides, the tetracyclines and trimethoprim. The shared presence of resistance classes alongside varying underlying ARG repertoires suggests that comparable resistance potential can arise through different genetic determinants. Isolation source was also not directly correlated with genomic relatedness, as was observed with the whole-genome analysis. One clinical and one environmental isolate were the closest genomic pair, and both the clinical and environmental groups contained ST69. The negative correlation between Mash distance and the resistome similarity almost reached statistical significance, suggesting that the number of ARGs shared by more closely related genomes was more and more similar, but did not necessarily form a genome-wide correlation. The plasmid replicon diversity was also highly diverse, and there was no correlation between replicon number and ARG burden. These observations suggest that other factors besides the clonal background affect the resistance composition.

The high level of ARG that was detected in environmental isolates is in agreement with genomic studies that have revealed that urban and aquatic environments can support a wide range of multidrug-resistant (MDR) *E. coli*

populations that carry clinically relevant ARGs (Tettey*et al*., 2024). Aquatic systems have also been demonstrated to be reservoirs of antimicrobial resistance and extraintestinal pathogenicity-associated characteristics in *E. coli* , further underscoring their significance in this regard (Gomi*et al*., 2017). Here, the cross-source genomic relatedness is also consistent with evidence from Thailand, where whole-genome sequencing revealed ESBL-producing *E. coli* from different epidemiological compartments that were genomically related, suggesting that ESBL-associated lineages can spread between compartments (Runcharoen*et al*., 2017). There is biological plausibility for such overlap, as *E. coli* has a wide diversity of genomic and ecological types, and a pathogenic classification is not sufficient to accurately describe the relationship of genomic lineages (Riley, 2020).

Environmental genomic studies have also shown that the capacity for resistance to antimicrobial agents can be merged with other traits that are beneficial for persistence and adaptation outside of the host, potentially contributing to the maintenance of antimicrobial-resistant populations in environmental settings, especially urban areas (Saini*et al*., 2024). Clinical populations are also diverse, as clinical strains of invasive *E. coli* have also been found to have significant inter-strain and inter-lineage differences, with genomic surveillance showing a high degree of variation in sequence types, antimicrobial resistance characteristics and clinically important lineages (Arconada Nuin*et al*., 2024). The identification of distinct multidrug-resistance profiles from the genomes of different clinical infections is similar to the recent comparative genome sequencing of strains of uropathogenic *E. coli* , in which unique combinations of acquired resistance genes were observed among the strains, despite their belonging to the same species (Manna*et al*., 2025). The sharing of plasmid replicons and the presence of source-restricted replicons are in line with the reports of clinically relevant multidrug-resistant and ESBL-producing *E. coli* in the environment and how mobile genetic elements can facilitate the dissemination of resistance across ecological niches (Teng*et al*., 2019).

The data from the combined resistome, MLST, Mash, and plasmid results showed that the two genetic populations of multidrug-resistant *E. coli* (clinical and environmental) should not be considered genetically distinct. The presence of shared sequence types, overlapping cross-source sequences and ARG and plasmid profiles suggests a related resistance landscape in which related bacterial backgrounds and mobile resistance determinants might exist in various environments. Environmental surveillance could therefore be a useful tool in addition to clinical AMR surveillance, especially when genomic methods have been applied to identify lineage-relatedness as opposed to similar bacteria due to acquired resistance elements. There is also a practical relevance, as there is no association between the number of plasmid replicons and ARG burden. Richness of replicons is not sufficient to quantify resistance potential, as genomes with relatively low numbers of replicon types can have large acquired resistomes. A better framework for comparative AMR surveillance is achieved by integrating ARG identity, plasmid background, sequence type and genome-wide relatedness.

There are some restrictions to be taken into account. The analysis was limited to only six genomes, which limits the statistical power, and the differences between the source levels should be viewed as patterns of the selected genomes and not as estimates of the population level. There was also variation in the quality of assembly of the isolates, which might account for the different numbers of fragmented genomic features detected. Plasmid replicon screening revealed signatures corresponding to plasmids but was not able to assign specific ARGs to each plasmid. Further studies are needed that involve larger clinical and environmental collections, sampled in similar geographic and temporal environments. The direct sequencing of long reads or genome assembly using hybrid reads would allow the precise determination of plasmid structures and ARG genomic environments, and would allow for larger comparative datasets to be used to determine if the cross-source sequence types and resistome patterns observed in this study were representative of the population.

5. CONCLUSION

The comparison of genomic sequences showed some differences but also overlapping antimicrobial resistance patterns of environmental and clinical isolates of *E. coli* . Resistance determinants and resistance classes were shared between clinical and environmental isolates, and there was a higher acquired ARG burden and more overall ARG diversity among environmental isolates. Whole-genome relatedness was not consistently observed based on the (clinical and environmental) source of isolation, and the detection of ST69 in both clinical and environmental genomes further confirmed that there is cross-source lineage overlap. Resistome similarity had only a medium negative correlation with genomic distance, suggesting that the acquired resistome composition was not only dependent on the level of clonally relatedness. Plasmid replicon profile varied, and there was no association between the number of replicons and ARG burden, with IncFIB(AP001918) being widely represented in the dataset. These results suggest that clinical and environmental MDR *E. coli* may share backgrounds and have different resistance repertoires. The findings provide a context for integrated genomic surveillance of clinical and environmental compartments to improve the understanding of the distribution of resistant lineages, acquired ARGs and mobile genetic elements.

REFERENCES

1. Ahmed, S. K., Hussein, S., Qurbani, K., Ibrahim, R. H., Fareeq, A., Mahmood, K. A., & Mohamed, M. G. (2024). Antimicrobial resistance: Impacts, challenges, and future prospects. *Journal of Medicine, Surgery, and Public Health*, *2*, 100081.

2. Al-Khalaifah, H., Rahman, M. H., Al-Surrayai, T., Al-Dhumair, A., & Al-Hasan, M. (2025). A One-Health perspective of antimicrobial resistance (AMR): human, animals and environmental health. *Life*, *15*(10), 1598.

3. Amann, S., Neef, K., & Kohl, S. (2019). Antimicrobial resistance (AMR). *European Journal of Hospital Pharmacy*, *26*(3), 175-177.

4. Arconada Nuin, E., Vilken, T., Xavier, B. B., Doua, J., Morrow, B., Geurtsen, J., ... & Van Puyvelde, S. (2024). A microbiological and genomic perspective of globally collected Escherichia coli from adults hospitalized with invasive *E. coli* disease. *Journal of Antimicrobial Chemotherapy*, *79*(9), 2142-2151.
5. Exner, M., Bhattacharya, S., Christiansen, B., Gebel, J., Goroncy-Bermes, P., Hartemann, P., ... & Trautmann, M. (2017). Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria?. *GMS hygiene and infection control*, *12*, Doc05.
6. Gomi, R., Matsuda, T., Matsumura, Y., Yamamoto, M., Tanaka, M., Ichiyama, S., & Yoneda, M. (2017). Whole-genome analysis of antimicrobial-resistant and extraintestinal pathogenic Escherichia coli in river water. *Applied and Environmental Microbiology*, *83*(5), e02703-16.
7. Hara, H., Yusaimi, Y. A., Zulkeflle, S. N. M., Sugiura, N., Iwamoto, K., Goto, M., Utsumi, M., Othman, N. B., & Zakaria, Z. (2018). Molecular characterization of multi-drug resistant *Escherichia coli* isolates from tropical environments in Southeast Asia. *The Journal of General and Applied Microbiology*, *64*(6), 284–292. https://doi.org/10.2323/jgam.2018.02.003
8. Ibrahim, D. R., Dodd, C. E., Stekel, D. J., Meshioye, R. T., Diggle, M., Lister, M., & Hobman, J. L. (2023). Multidrug-resistant ESBL-producing *E. coli* in clinical samples from the UK. *Antibiotics*, *12*(1), 169.
9. Macesic, N., Uhlemann, A. C., & Peleg, A. Y. (2025). Multidrug-resistant Gram-negative bacterial infections. *The Lancet*, *405*(10474), 257-272.
10. Manna, A., Ramadoss, R., K, V., S, S., R, R., Ramasamy, D., & Lakshmanan, D. (2025). Whole-genome sequencing and comparative genome analysis of three multidrug-resistant clinical uropathogenic Escherichia coli isolates from India. *Journal of Applied Microbiology*, *136*(9), lxaf228.
11. Matsumura, Y., Yamamoto, M., Gomi, R., Tsuchido, Y., Shinohara, K., Noguchi, T., & Nagao, M. (2025). Integrating whole-genome sequencing into antimicrobial resistance surveillance: methodologies, challenges, and perspectives. *Clinical Microbiology Reviews*, *38*(4), e00140-22.
12. Meradji, S., Basher, N. S., Sassi, A., Ibrahim, N. A., Idres, T., & Touati, A. (2025). The role of water as a reservoir for antibiotic-resistant bacteria. *Antibiotics*, *14*(8), 763.
13. Moo, C. L., Yang, S. K., Yusoff, K., Ajat, M., Thomas, W., Abushelaibi, A., ... & Lai, K. S. (2020). Mechanisms of antimicrobial resistance (AMR) and alternative approaches to overcome AMR. *Current drug discovery technologies*, *17*(4), 430-447.
14. Neumann, B., Rackwitz, W., Hunfeld, K. P., Fuchs, S., Werner, G., & Pfeifer, Y. (2020). Genome sequences of two clinical Escherichia coli isolates harboring the novel colistin-resistance gene variants mcr-1.26 and mcr-1.27. *Gut Pathogens*, *12*(1), 40.
15. Odonkor, S. T., & Addo, K. K. (2018). Prevalence of multidrug-resistant Escherichia coli isolated from drinking water sources. *International journal of microbiology*, *2018*(1), 7204013.
16. Read, D. S., Gweon, H. S., Bowes, M. J., Anjum, M. F., Crook, D. W., Chau, K. K., ... & Stoesser, N. (2024). Dissemination and persistence of antimicrobial resistance (AMR) along the wastewater-river continuum. *Water research*, *264*, 122204.
17. Riley, L. W. (2020). Distinguishing pathovars from nonpathovars: Escherichia coli. *Microbiology spectrum*, *8*(4), 10-1128.
18. Runcharoen, C., Raven, K. E., Reuter, S., Kallonen, T., Paksanont, S., Thammachote, J., ... & Chantratita, N. (2017). Whole genome sequencing of ESBL-producing Escherichia coli isolated from patients, farm waste and canals in Thailand. *Genome medicine*, *9*(1), 81.
19. Saini, P., Bandsode, V., Singh, A., Mendem, S. K., Semmler, T., Alam, M., & Ahmed, N. (2024). Genomic insights into virulence, antimicrobial resistance, and adaptation acumen of Escherichia coli isolated from an urban environment. *MBio*, *15*(3), e03545-23.
20. Tang, K. W. K., Millar, B. C., & Moore, J. E. (2023). Antimicrobial resistance (AMR). *British journal of biomedical science*, *80*, 11387.
21. Teng, L., Lee, S., Ginn, A., Markland, S. M., Mir, R. A., DiLorenzo, N., ... & Jeong, K. C. (2019). Genomic comparison reveals natural occurrence of clinically relevant multidrug-resistant extended-spectrum-β-lactamase-producing Escherichia coli strains. *Applied and environmental microbiology*, *85*(13), e03030-18.
22. Tettey, R., Egyir, B., Tettey, P., Arko-Mensah, J., Addo, S. O., Owusu-Nyantakyi, C., ... & Fobil, J. (2024). Genomic analysis of multidrug-resistant Escherichia coli from Urban Environmental water sources in Accra, Ghana, Provides Insights into public health implications. *PLoS One*, *19*(5), e0301531.