Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated Phenotypic and Whole-Genome Characterization of a Uropathogenic Escherichia coli ST127, O6:H31 Isolate from Bangladesh

Md. Asif Karim
Jagannath University
Sohana Parveen Meem
Jagannath University
Nizam Uddin
Bangladesh Council of Scientific and Industrial Research (BCSIR)
Tanjina Akhtar Banu
Bangladesh Council of Scientific and Industrial Research (BCSIR)
Mousona Islam
Bangladesh Council of Scientific and Industrial Research (BCSIR)
Barna Goswami
Bangladesh Council of Scientific and Industrial Research (BCSIR)
Sanjana Fatema Chowdhury
Bangladesh Council of Scientific and Industrial Research (BCSIR)
Showti Raheel Naser
Bangladesh Council of Scientific and Industrial Research (BCSIR)
Sadia Manzoor
University of the Punjab
Jonas Ivan Nobre Oliveira
Federal University of Rio Grande do Norte
Shamima Begum
Jagannath University
Shahina Akter
shupty2010@gmail.com

Bangladesh Council of Scientific and Industrial Research (BCSIR)

Article

Keywords: Antimicrobial resistance, Multidrug resistance, Escherichia coli O6:H31, BCSIR SA 01, Urinary tract infection, uropathogens, whole-genome sequencing

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11428884/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrated Phenotypic and Whole-Genome Characterization of a Uropathogenic *Escherichia coli* ST127, O6:H31 Isolate from Bangladesh

Md. Asif Karim[Φ2], Sohana Parveen Meem[Φ2], Nizam Uddin[1], Tanjina Akhtar Banu [1], Mousona Islam[1], Barna Goswami[1], Sanjana Fatema Chowdhury[1], Showti Raheel Naser[1], Sadia Manzoor[3], Jonas Ivan Nobre Oliveira[4] Shamima Begum[2], Shahina Akter[1]*

[1] Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka-1205, Bangladesh

[2] Department of Microbiology, Jagannath University, Dhaka 1100, Bangladesh

[3]Centre for Applied Molecular Biology (CAMB), University of the Punjab, Quaid-e-Azam Campus, Lahore, Pakistan

[4] Department of Biophysics and Pharmacology, Federal University of Rio Grande do Norte, 59072-970, Natal/RN, Brazil

*Corresponding author Email: Dr. Shahina Akter shupty2010@gmail.com

Φ Equal contribution as first author

Abstract

Background: Urinary tract infections caused by uropathogenic *Escherichia coli* (UPEC) are increasingly complicated by antimicrobial resistance and diverse virulence-associated genomic features. Whole-genome sequencing (WGS) provides a comprehensive approach for characterizing these determinants within individual clinical isolates.

Methods: In this study, a clinical *E. coli* isolate, designated BCSIR SA 01, was characterized using antimicrobial susceptibility testing and whole-genome sequencing. The draft genome was assembled using SPAdes and assessed using QUAST and CheckM2, followed by genome annotation, genomic typing, antimicrobial resistance and virulence profiling, genomic island prediction, mobile genetic element and prophage analysis, CRISPR-associated feature detection, secondary metabolite prediction, and in silico pathogenicity assessment.

Results: The genome was 5,210,642 bp in length and comprised 234 contigs, with 50.24% G+C content, 100.00% estimated completeness, and 1.39% contamination. The isolate was classified as *E. coli* phylogroup B2, serotype O6:H31, MLST ST127, and FimH type fimH180, with a FastANI value of 98.61%. Phenotypically, the isolate was resistant to ampicillin, ceftazidime, ceftriaxone, and nalidixic acid, while remaining susceptible to amikacin and gentamicin. CARD/RGI analysis identified an EC-5 β-lactamase homolog, *tet(B)*, *tetB(P)*-like, *gyrA* S83L, and additional efflux- and regulatory-associated resistance models. The genome also contained multiple virulence-associated determinants, including *hlyA-hlyD*, *cnf1*, *pap*, *sfa*, *fim*, *iroBCDEN*, *ybt/fyuA/irp*, *chu*,

and *kps* genes. Several predicted genomic islands contained virulence-, resistance-, siderophore-, capsule-, and mobile-element-associated features.

Conclusion: BCSIR SA 01 exhibited a complex combination of antimicrobial resistance and virulence-associated genomic characteristics within an ST127/O6:H31/B2 background. These findings demonstrate the utility of integrated phenotypic and WGS-based analysis for comprehensive characterization of clinically relevant UPEC isolates.

Keywords: Antimicrobial resistance; Multidrug resistance; *Escherichia coli* O6:H31; BCSIR SA 01; Urinary tract infection, uropathogens, whole-genome sequencing

1. Introduction

Urinary tract infections (UTIs) are among the most common bacterial infections affecting humans and represent a substantial global health burden. (Flores-Mireles *et al*., 2015). UTIs encompass a range of clinical conditions involving the urinary tract, including uncomplicated cystitis and more severe infections that may progress to upper urinary tract involvement. (Flores-Mireles *et al*., 2015). Among the bacterial pathogens responsible for UTIs, *Escherichia coli* is the predominant cause of community-acquired uncomplicated urinary infections. (Flores-Mireles *et al*., 2015; Wiles, Kulesus and Mulvey, 2008). Uropathogenic *E. coli* (UPEC) represents a pathogenic group of *E. coli* that possesses genetic and phenotypic traits enabling colonization, persistence, and disease within the urinary tract. (Wiles, Kulesus and Mulvey, 2008; Subashchandrabose and Mobley, 2015). Successful UPEC infection depends on a diverse repertoire of virulence and fitness determinants that facilitate bacterial attachment, invasion, nutrient acquisition, persistence, and adaptation to host defenses. (Subashchandrabose and Mobley, 2015). Adhesive structures such as type 1 fimbriae and P fimbriae contribute to attachment and colonization of urinary tract tissues, whereas iron-acquisition systems enable bacterial growth under the iron-limited conditions encountered within the host. (Wiles, Kulesus and Mulvey, 2008; Subashchandrabose and Mobley, 2015). Other virulence-associated determinants, including hemolysin, cytotoxic necrotizing factor 1, capsules, toxins, autotransporters, flagellar systems, and secretion systems, further contribute to the biological diversity and pathogenic potential of UPEC. (Johnson, 1991; Wiles, Kulesus and Mulvey, 2008; Subashchandrabose and Mobley, 2015). The distribution of these determinants is highly variable among UPEC strains, reflecting extensive genetic diversity within the UPEC population. (Wiles, Kulesus and Mulvey, 2008; Subashchandrabose and Mobley, 2015). Many virulence-associated genes are carried within the accessory genome, including pathogenicity islands and other horizontally acquired genomic regions that can contain clusters of functionally related genes. (Hacker and Kaper, 2000). Pathogenicity islands are recognized as important contributors to bacterial genome plasticity and can facilitate the acquisition and coordinated expression of virulence-associated functions. (Hacker and Kaper, 2000). In parallel with virulence, antimicrobial resistance (AMR) has become a major challenge in the treatment and control of

bacterial urinary tract infections. (Pitout, 2012; World Health Organization, 2026). The World Health Organization reported that approximately one in six laboratory-confirmed bacterial infections worldwide were resistant to antibiotics in 2023, highlighting the continuing global expansion of antimicrobial resistance. (World Health Organization, 2026). Resistance in *E. coli* may result from diverse mechanisms, including enzymatic inactivation of antimicrobial agents, modification of antibiotic targets, reduced intracellular drug accumulation, and active efflux. (Pitout, 2012; McArthur *et al*., 2013). β-lactamase production is an important mechanism of resistance in extraintestinal pathogenic *E. coli*, while mutations in antimicrobial targets and acquisition of resistance-associated genes can additionally contribute to multidrug-resistant phenotypes. (Pitout, 2012; McArthur *et al*., 2013). The coexistence of resistance and virulence-associated determinants may generate bacterial strains with complex clinical and genomic profiles, although the relationship between antimicrobial resistance and virulence varies according to bacterial species, genetic mechanism, ecological niche, and host environment. (Cepas and Soto, 2020). This combination of phenotypic resistance and diverse virulence-associated features makes comprehensive characterization of clinical *E. coli* isolates particularly important. (Subashchandrabose and Mobley, 2015; Cepas and Soto, 2020). The situation is particularly relevant in Bangladesh, where clinical studies have documented substantial antimicrobial resistance among uropathogenic *E. coli* isolates. (Islam *et al*., 2022). A recent study from a tertiary-care hospital in Dhaka reported that *E. coli* was the most frequently recovered Gram-negative uropathogen and that a large proportion of *E. coli* isolates exhibited multidrug resistance, with particularly high resistance observed against several commonly used antimicrobial agents. (Chowdhury *et al*., 2024). Molecular investigations in Bangladesh have also identified clinically important resistance and virulence determinants among ESBL-producing *E. coli*, together with mobile genetic elements and plasmid-associated genomic features. (Moniruzzaman *et al*., 2023). More recently, whole-genome-based surveillance of community-acquired ESBL-producing *E. coli* in Bangladesh has demonstrated substantial genomic diversity and highlighted the importance of genomic approaches for understanding the origins and dissemination of resistant urinary isolates. (Amin *et al*., 2026). Whole-genome sequencing (WGS) provides nucleotide-level information that can simultaneously support species identification, strain typing, antimicrobial resistance profiling, virulence characterization, and analysis of accessory genomic features. (Vanstokstraeten *et al*., 2023). Compared with phenotypic antimicrobial susceptibility testing alone, WGS can reveal both acquired resistance genes and chromosomal mutations associated with resistance, thereby providing a genomic perspective on observed antimicrobial phenotypes. (Vanstokstraeten *et al*., 2023). WGS-based characterization can additionally identify phylogenetic relationships, serotypes, sequence types, fimH alleles, genomic islands, prophage-associated sequences, and other mobile genetic elements that may contribute to the genomic diversity of UPEC. (Jolley, Bray and Maiden, 2018; Joensen et al., 2015; Roer et al., 2017; Jain et al., 2018; Brown et al., 2022). Sequence type ST127 is one of the recognized UPEC lineages and has been identified among community-associated urinary isolates, including isolates carrying ESBL-associated resistance determinants (Cave et al., 2022). The ST127 lineage has also been reported in association with the

O6:H31 serotype, illustrating the relationship between specific clonal backgrounds and UPEC virulence-associated genomic profiles. (Johnson et al*.,* 2008). Despite the increasing availability of phenotypic and targeted molecular data, integrated characterization of individual clinical *E. coli* isolates can provide additional insight by linking antimicrobial resistance, virulence, phylogenetic identity, and accessory-genome architecture within the same genomic framework. (Vanstokstraeten et al., 2023; Amin et al., 2026). Accordingly, comprehensive genome analysis is particularly useful for evaluating the biological features of clinically relevant isolates and for identifying genomic determinants that may contribute to their resistance and pathogenic potential. (Subashchandrabose and Mobley, 2015; Vanstokstraeten et al., 2023).

In this study, we performed an integrated phenotypic and whole-genome characterization of the clinical *E. coli* isolate BCSIR SA 01, combining antimicrobial susceptibility testing with genome assembly, species confirmation, genomic typing, antimicrobial resistance analysis, virulence profiling, genomic island prediction, mobile genetic element and prophage analysis, CRISPR-associated feature detection, functional annotation, secondary metabolite biosynthetic potential assessment, and In silico pathogenicity prediction.

The study aimed to define the genomic characteristics of the isolate and to examine the relationship between its observed antimicrobial susceptibility phenotype and the resistance-associated genomic determinants identified through whole-genome analysis.

2. Materials and Methods

2.1 Clinical Specimen Collection

A midstream clean-catch urine specimen was collected in a sterile urine collection tube from a patient presenting with symptoms of acute uncomplicated urinary tract infection (uUTI). (Leber, 2016; Tajbakhsh, Tajbakhsh and Khamesipour, 2015). The specimen was collected in the morning and transported to the laboratory for processing within 1 h of collection. (Leber, 2016; Tajbakhsh, Tajbakhsh and Khamesipour, 2015). The urine specimen was subsequently used for bacterial isolation, antimicrobial susceptibility testing, and genomic characterization. (Leber, 2016; Rahman et al., 2022).

2.2 Bacterial Isolation and Phenotypic Identification

For bacterial isolation, the urine specimen was inoculated into nutrient broth and subsequently cultured on MacConkey agar and eosin methylene blue (EMB) agar. (Tajbakhsh, Tajbakhsh and Khamesipour, 2015; Rahman et al., 2022). The inoculated media were incubated aerobically at 37°C for 24 h to facilitate bacterial growth. (Tajbakhsh, Tajbakhsh and Khamesipour, 2015). Presumptive identification of *E. coli* was based on characteristic colony morphology on MacConkey agar and the typical metallic-green sheen observed on EMB agar. (Tajbakhsh, Tajbakhsh and Khamesipour, 2015; Rahman et al., 2022). The isolate selected for whole-genome sequencing and subsequent genomic analyses was designated *E. coli* BCSIR SA 01.

2.3 Antimicrobial Susceptibility Testing

Antimicrobial susceptibility testing (AST) was performed using the Kirby–Bauer disk diffusion method on Mueller–Hinton agar. (Clinical and Laboratory Standards Institute [CLSI], 2024; Rahman et al., 2022). A bacterial suspension equivalent to a 0.5 McFarland turbidity standard was prepared to standardize the inoculum prior to plate inoculation. (CLSI, 2024). The inoculated Mueller–Hinton agar plates containing antimicrobial disks were incubated at 37°C for 24 h under aerobic conditions. (CLSI, 2024). The antimicrobial agents tested included ampicillin (10 µg), ceftazidime (30 µg), ceftriaxone (30 µg), nalidixic acid (30 µg), amikacin (30 µg), and gentamicin (10 µg), (CLSI, 2024). After incubation, inhibition-zone diameters were measured in millimeters and interpreted according to the applicable CLSI antimicrobial susceptibility criteria (CLSI, 2024).

2.4 Genomic DNA Extraction and Quantification

Genomic DNA was extracted from the bacterial pellet using the QIAamp DNA Mini Kit according to the manufacturer's spin-column protocol for bacterial genomic DNA extraction. (QIAGEN, 2020). The extracted genomic DNA was assessed for concentration and purity before preparation of whole-genome sequencing libraries. (QIAGEN, 2020). Double-stranded DNA concentration and yield were quantified using a Qubit 4.0 Fluorometer with a dsDNA High-Sensitivity assay, while DNA purity was assessed using a NanoDrop 2000 spectrophotometer. (Thermo Fisher Scientific, 2019; Thermo Fisher Scientific, 2009; Rahman et al., 2022).

2.5 Whole-Genome Sequencing and Library Preparation

Sequencing libraries were constructed using the Illumina DNA Prep Kit following the manufacturer's workflow for DNA fragmentation and indexed library preparation. (Illumina Inc., 2020; Illumina, 2026). The prepared libraries were sequenced on the Illumina MiniSeq platform at the designated sequencing facility of the Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh. (Illumina Inc., 2018; Illumina, 2021; Akter et al., 2023; Chowdhury et al., 2026; Al Sium et al., 2025; Mohammad et al., 2025). Paired-end sequencing was performed to generate reads for downstream de novo genome assembly and genomic characterization (Illumina Inc., 2018; Illumina, 2021). The raw sequencing data for BCSIR SA 01 were submitted to the NCBI under BioProject accession number PRJNA1537254, BioSample accession number SAMN63953655, and Sequence Read Archive (SRA) accession numbers SRX35532930 (Experiment) and SRR41005028 (Run). (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1537254)

2.6 De novo Genome Assembly and Post-assembly Contig Filtering

The paired-end sequencing reads were de novo assembled using the SPAdes genome assembler implemented through the Galaxy platform, with the Galaxy tool version reported as 4.3.0+galaxy2. (Bankevich et al., 2012; Afgan et al., 2018). SPAdes perform de novo genome assembly using a de Bruijn graph-based approach and is widely used for bacterial whole-genome assembly from short-read sequencing data. (Bankevich et al., 2012). The sequencing reads were first assembled

into contigs using SPAdes before any length-based filtering was performed. (Bankevich et al., 2012). Following assembly, contigs shorter than 1,000 bp were removed as a post-assembly size-filtering step, and the remaining contigs were retained as the final draft assembly FASTA file for downstream analyses. (Bankevich et al., 2012).

2.7 Genome Assembly Quality Assessment

The quality of the final draft genome assembly was assessed using QUAST to obtain standard assembly statistics, including total assembly length, number of contigs, largest contig, G+C content, and the number of ambiguous bases. (Gurevich et al., 2013). Genome completeness and contamination were assessed using CheckM2 to evaluate the overall quality of the microbial genome assembly. (Chklovski et al., 2023). The final genome was classified as a draft assembly because the genome sequence remained distributed across multiple contigs rather than a single closed chromosome. (Gurevich et al., 2013).

2.8 Genome Annotation and Functional Characterization

The final draft genome assembly was annotated using Bakta with the full database version 6.0 to identify coding sequences and other genomic features. (Schwengers et al., 2021). A complementary genome annotation was performed using the RASTtk-based RAST framework for subsystem-oriented functional characterization. (Brettin et al., 2015). Clusters of Orthologous Groups (COG)-based functional classification was performed using the BV-BRC platform to categorize predicted proteins according to conserved functional groups. (Tatusov et al., 2000; Olson et al., 2023). KEGG-based functional annotation was performed using BlastKOALA in conjunction with the BV-BRC workflow to assign predicted proteins to KEGG Orthology groups and related functional pathways. (Kanehisa, Sato and Morishima, 2016; Kanehisa et al., 2025; Kanehisa and Goto, 2000; Olson et al., 2023).

2.9 Genomic Typing and Strain Characterization

Species- and strain-level characterization was performed using multiple complementary In silico approaches applied to the final draft genome assembly. (Jolley, Bray and Maiden, 2018; Jain et al., 2018; Meier-Kolthoff and Göker, 2019). Multilocus sequence typing (MLST) was performed using the PubMLST/BIGSdb framework to determine the sequence type based on the established allele scheme. (Jolley, Bray and Maiden, 2018). In silico Clermont phylogrouping was performed to classify the isolate into the established *E. coli* phylogenetic groups. (Clermont et al., 2013). In silico O- and H-antigen serotyping was performed using the SerotypeFinder approach for whole-genome sequence data. (Joensen et al., 2015). CH typing was performed using the *fumC* and *fimH* allele-based scheme to provide additional discriminatory typing information for extraintestinal pathogenic *E. coli*. (Weissman et al., 2012; Roer et al., 2018). FimH subtyping was performed using the genome-based FimTyper approach to determine the *fimH* allele type. (Roer et al., 2017). Average nucleotide identity (ANI) was calculated using FastANI to assess whole-genome nucleotide similarity and support species-level genome classification. (Jain et al., 2018). Genome-

based taxonomic placement and phylogenomic analysis were performed using the Type (Strain) Genome Server (TYGS). (Meier-Kolthoff and Göker, 2019).

2.10 Antimicrobial Resistance Determinant Analysis

Antimicrobial resistance-associated genomic determinants were identified using the Comprehensive Antibiotic Resistance Database (CARD) together with the Resistance Gene Identifier (RGI) analysis framework. (McArthur et al., 2013; Alcock et al., 2023). The analysis was performed using the latest CARD/RGI configuration available in the analysis environment with the default RGI settings. (Alcock et al., 2023). No additional manually defined identity or reference-coverage thresholds were applied during the analysis. (Alcock et al., 2023). Detected resistance-associated features were classified according to the corresponding CARD model type and predicted resistance mechanism, including protein homolog, protein variant, efflux-associated, and regulatory resistance models where applicable. (McArthur et al., 2013; Alcock et al., 2023).

2.11 Virulence-associated Gene Analysis

Virulence-associated genomic determinants were identified by analysing the final draft genome against the Virulence Factor Database (VFDB) using a Galaxy-based analysis workflow. (Liu et al., 2022; Afgan et al., 2018). The analysis was used to identify genomic determinants associated with major virulence categories, including toxins, adhesion, iron acquisition, immune evasion, invasion, motility, chemotaxis, and secretion systems. (Liu et al., 2022).

2.12 Genomic Island Prediction

Genomic island prediction was performed using IslandViewer 4 with the final draft genome FASTA file as the input sequence. (Bertelli et al., 2017). The analysis was performed using the default settings provided by the IslandViewer 4 platform. (Bertelli et al., 2017). Predicted genomic island regions and their coordinates were retained as reported in the IslandViewer 4 output without manual modification. (Bertelli et al., 2017). The predicted regions were subsequently examined for genes and features associated with virulence, antimicrobial resistance, capsule biosynthesis, siderophore acquisition, adhesion, secretion, and mobile genetic elements. (Bertelli et al., 2017).

2.13 Prophage, Viral and Mobile Genetic Element Analysis

Putative prophage-associated regions were identified using PHASTEST to detect and characterize prophage-like sequences within the draft genome. (Arndt et al., 2019). PHIGARO was additionally used to identify bacteriophage-associated regions as a complementary prophage prediction approach. (Starikova et al., 2020). VirSorter was used to screen the genome for sequence signatures associated with viral genomic regions. (Roux et al., 2015). Mobile genetic element-associated proteins were identified using mobileOG-db, a curated database of protein families involved in bacterial mobile genetic element biology. (Brown et al., 2022). PHASTEST, PHIGARO, VirSorter, and mobileOG-db analyses were performed using the default settings of the respective workflows. (Arndt et al., 2019; Starikova et al., 2020; Roux et al., 2015; Brown et al., 2022).

2.14 CRISPR-associated Sequence Analysis

CRISPR-associated genomic features were identified using the CRISPR/Cas Finder tool implemented in the Proksee platform (Proksee tool version 1.1.0; CRISPRCasFinder version 4.2.20). (Proksee, 2026). The tool was used to identify putative CRISPR arrays and their associated Cas proteins within the draft genome assembly. (Proksee, 2026). The analysis was performed using the default settings available in the Proksee CRISPR/Cas Finder workflow. (Proksee, 2026). The underlying CRISPRCasFinder methodology integrates CRISPR-array detection with identification of associated Cas proteins for characterization of CRISPR–Cas systems. (Couvin et al., 2018).

2.15 Secondary Metabolite Biosynthetic Gene Cluster Analysis

The final bacterial genome FASTA file was submitted to antiSMASH for prediction of secondary metabolite and specialized metabolite biosynthetic gene clusters. (Blin et al., 2023). The analysis was conducted using the default antiSMASH settings, and the predicted biosynthetic gene cluster output was used for downstream genomic interpretation. (Blin et al., 2023).

2.16 In silico Pathogenicity Prediction

The pathogenic potential of the isolate was evaluated using PathogenFinder 2, version 0.6.0, with the default analysis settings. (Cosentino et al., 2013). PathogenFinder was used to provide a computational prediction of the association of the genome with human pathogenicity based on genomic features. (Cosentino et al., 2013). The resulting output was interpreted as an in-silico pathogenicity prediction and not as experimental evidence of pathogenicity or virulence in a host. (Cosentino et al., 2013).

2.17 Integration of Phenotypic and Genomic Resistance Data

The phenotypic antimicrobial susceptibility profile was compared descriptively with the resistance-associated genomic determinants identified by CARD/RGI. (McArthur et al., 2013; Alcock et al., 2023). Genotype–phenotype relationships were interpreted as associations between genomic findings and observed antimicrobial susceptibility phenotypes rather than direct experimental evidence of causality. (CLSI, 2024; Alcock et al., 2023). Resistance-associated variants identified through models developed for organisms other than *E. coli* were interpreted cautiously and were not considered independently validated without organism-specific sequence confirmation. (Alcock et al., 2023).

2.18 Data Availability

The raw sequencing data for BCSIR SA 01 were submitted to the NCBI under BioProject accession number PRJNA1537254, BioSample accession number SAMN63953655, and Sequence Read Archive (SRA) accession numbers SRX35532930 (Experiment) and SRR41005028 (Run) (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1537254).

2.19 Ethics Approval and Informed Consent

The study received ethical approval from the Ethical Review Committee of Jagannath University, Dhaka, Bangladesh (Approval Code: JnU/ERC/2026/64) on 16 September 2026. All methods were performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from the participant before specimen enrollment and use of the clinical specimen for research purposes. Clinical information used for the study was handled in an anonymized manner to protect participant confidentiality.

3. Results

3.1. Clinical and isolate characteristics

E. coli BCSIR SA 01 was recovered from a 27-year-old female presenting with acute dysuria and increased urinary frequency. The episode was classified as community-acquired uncomplicated UTI, and a midstream clean-catch urine specimen was collected. The subject had a history of incomplete antibiotic self-medication before sampling. Written informed consent was obtained.

3.2. Genome assembly and quality assessment

Table 1. Genome assembly and quality metrics of *Escherichia coli* BCSIR SA 01

Metric	Value
Total genome size	5,210,642 bp
Number of contigs	234
Largest contig	271,365 bp
G+C content	50.24%
Ambiguous bases	0
CheckM2 completeness	100.00%
CheckM2 contamination	1.39%
Coding density	87.7%
Predicted CDS	5,006
Average gene length	304.81 bp

Circular comparative genomic analysis of *Escherichia coli* BCSIR SA 01. The genomic architecture and structural conservation of *E. coli* BCSIR SA 01 were visualized using a circular comparative genomic map aligned against closely related reference genomes. As depicted in the figure, the outermost dark blue ring represents the reference genome coordinates, while the inner concentric rings display sequence similarity alignments (likely based on BLASTn) between the reference and the query genome. The widespread presence of colored alignment rings

across the majority of the circular map indicates a high degree of sequence conservation and synteny, confirming the close evolutionary relationship of strain BCSIR SA 01 to the reference strains. However, the disruption of these rings and the presence of gaps indicate regions of genomic divergence, which may harbor strain-specific insertions, deletions, or horizontally acquired elements such as prophages or genomic islands. The innermost tracks, displaying GC content (black jagged line) and GC skew (purple and tan shading), reveal localized fluctuations. These shifts in nucleotide composition often correspond to the boundaries of mobile genetic elements or horizontally transferred regions, further highlighting the genomic plasticity of the strain. Collectively, this comparative map provides a high-resolution visual overview of the genomic similarities and unique structural variations that define *E. coli* BCSIR SA 01 (Figure 1).

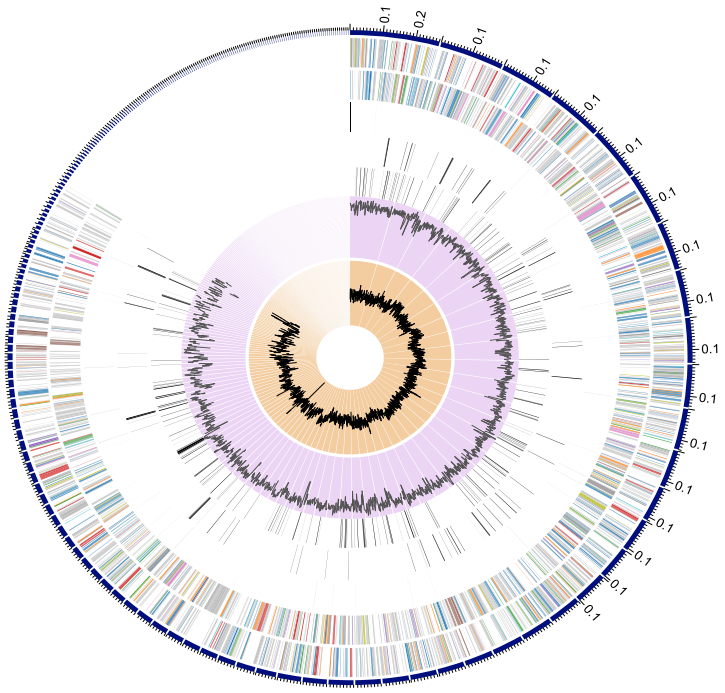


Figure 1: Circular comparative genomic analysis of *E. coli* BCSIR SA 01 against reference genomes. The circular map illustrates the genomic architecture and sequence conservation of the draft genome. The outer dark blue ring denotes the reference sequence(s), while the inner concentric rings represent the query genome(s) and sequence similarity alignments. The inner track displays the GC content (black line) and GC skew (purple/green shading).

3.3. Species identification and genomic typing

Table 2. *In silico* genomic typing of *E. coli* BCSIR SA 01

Typing / classification method	**Result**
Species identification	*Escherichia coli*

Clermont phylogroup	B2
Serotype	O6:H31
MLST sequence type	ST127
FimH type	fimH180
FastANI	98.61%

3.4. Phylogenomic placement

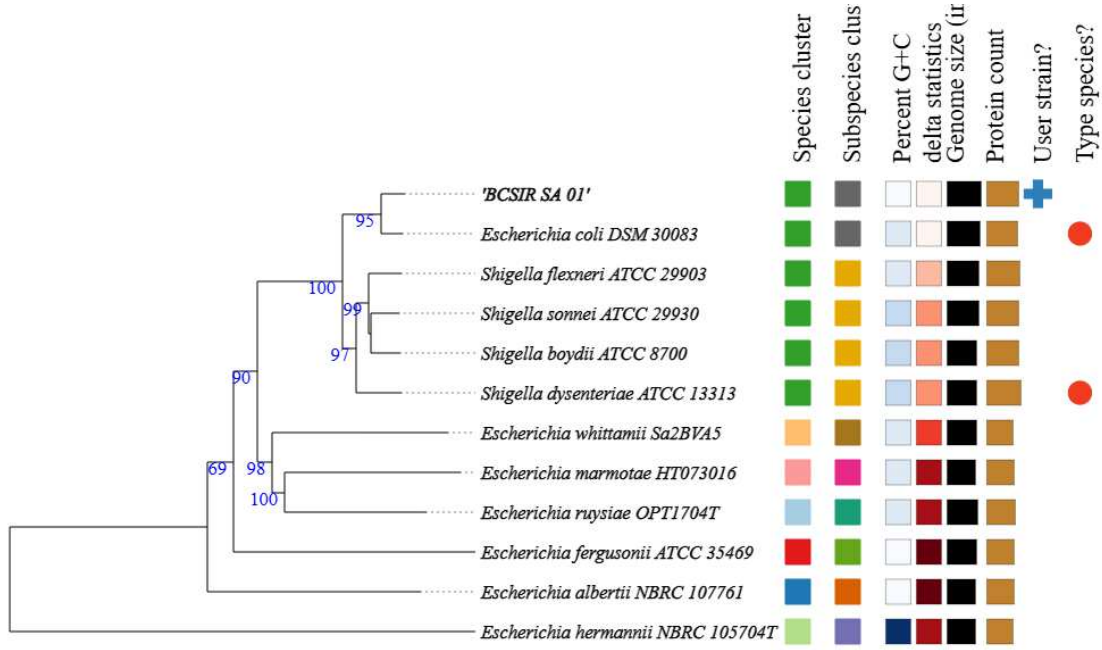
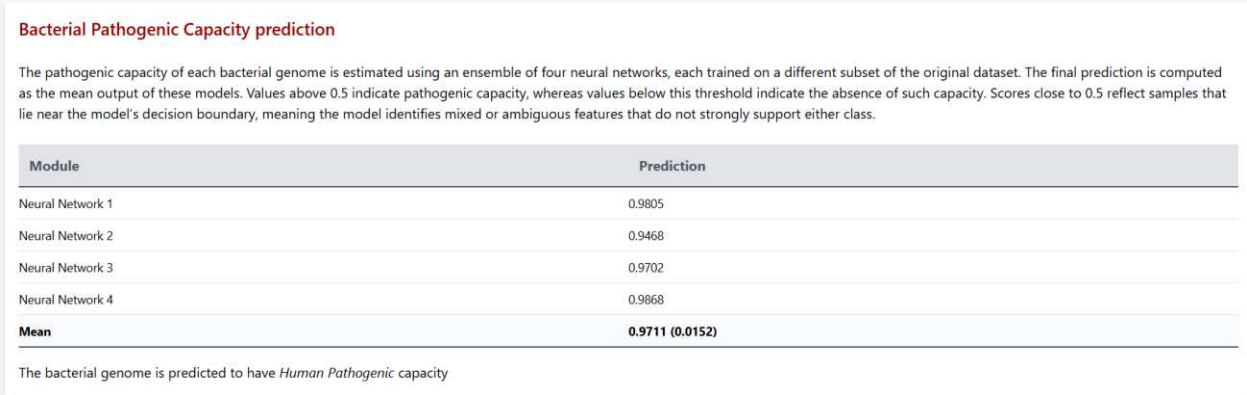


Figure 2: Phylogram (TYGS). TYGS-based phylogenomic analysis placed BCSIR SA 01 within the *E. coli* species cluster, supporting the genome-based species assignment.

Figure 2 presents a phylogenomic phylogram generated using the Type (Strain) Genome Server (TYGS), which illustrates the evolutionary placement of strain BCSIR SA 01 relative to established type strains. The analysis places BCSIR SA 01 firmly within the Escherichia coli species cluster, indicating a close genomic relationship with recognized E. coli type strains. This TYGS-based phylogenomic result is consistent with, and independently corroborates, the genome-based species assignment of the isolate, thereby providing robust support for its taxonomic classification as a member of the *E. coli* species complex.

3.5. *In silico* pathogenicity prediction

Figure 3A illustrates the *In silico* prediction of the pathogenic potential of the sequenced bacterial genome using the PathogenFinder tool. The software employs an ensemble of four independently trained neural networks to evaluate the presence of pathogenic features, generating a probability score between 0 and 1. As shown in the table, all four neural network modules returned high prediction scores, ranging from 0.9468 to 0.9868. The final consensus prediction is calculated as the mean of these outputs, yielding a score of 0.9711 (SD = 0.0152). Because this value significantly exceeds the 0.5 decision boundary threshold, the model confidently predicts that the bacterial genome possesses human pathogenic capacity, indicating that the strain harbors genetic features strongly associated with human virulence.



Bacterial Pathogenic Capacity prediction

The pathogenic capacity of each bacterial genome is estimated using an ensemble of four neural networks, each trained on a different subset of the original dataset. The final prediction is computed as the mean output of these models. Values above 0.5 indicate pathogenic capacity, whereas values below this threshold indicate the absence of such capacity. Scores close to 0.5 reflect samples that lie near the model's decision boundary, meaning the model identifies mixed or ambiguous features that do not strongly support either class.

Module	Prediction
Neural Network 1	0.9805
Neural Network 2	0.9468
Neural Network 3	0.9702
Neural Network 4	0.9868
Mean	**0.9711 (0.0152)**

The bacterial genome is predicted to have *Human Pathogenic* capacity

Figure 3A: *In silico* pathogenicity prediction using PathogenFinder. The pathogenic capacity of the bacterial genome was estimated using an ensemble of four neural networks. The table displays the prediction scores for each individual network module and the final mean score. A mean score above the 0.5 threshold indicates a prediction of human pathogenic capacity.

Figure 3B presents a two-dimensional UMAP (Uniform Manifold Approximation and Projection) visualization of the bacterial pathogenic landscape, mapping the genomic position of strain BCSIR SA 01 relative to a curated reference database of known pathogens. In the plot, red dots represent the reference pathogen genomes, while the yellow diamond marks the position of the query sequence (BCSIR SA 01). The analysis reveals that the genome of BCSIR SA 01 clusters tightly within a dense region of known pathogenic sequences, demonstrating high genomic similarity and shared features with established pathogens. Furthermore, the immediate proximity of the query to the "Closer Pathogens" markers (indicated by grey crosses) suggests that its genomic profile is highly comparable to the most closely related pathogenic strains in the dataset, reinforcing the *in silico* prediction of its pathogenic capacity.

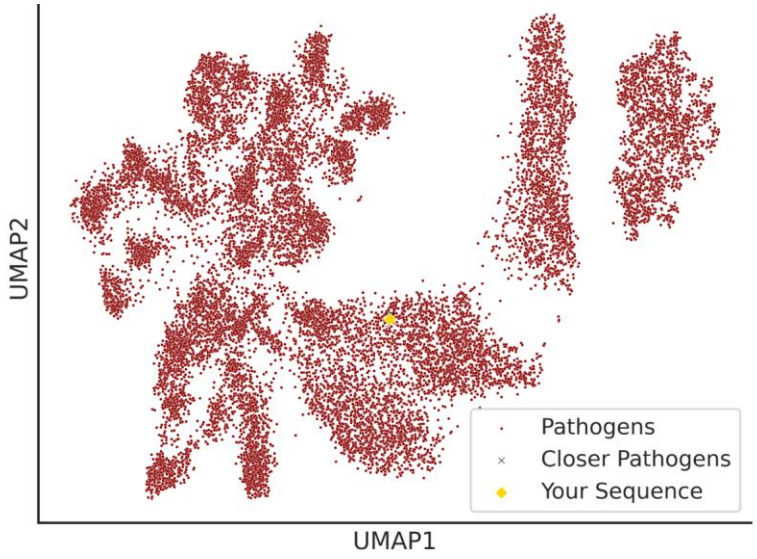


Figure 3B: Bacterial pathogenic landscape showing the genomic position of BCSIR SA 01 relative to reference pathogenic genomes.

3.6. Antimicrobial resistance-associated determinants

Table 3: Principal antimicrobial resistance-associated determinants identified in *Escherichia coli* BCSIR SA 01 by CARD/RGI analysis

Resistance-associated determinant	RGI detection/model	Identity (%)	Reference coverage (%)	Resistance association/interpretation
EC-5 β-lactamase	Strict; protein homolog model	99.73	100	β-lactam antibiotic inactivation; detected as an EC-5 β-lactamase homolog.
tet(B)	Strict; protein homolog model	99.25	100	MFS-mediated tetracycline efflux; strong resistance-associated determinant.
tetB(P)-like	Strict; protein homolog model	98.92	85.28	Tetracycline resistance-associated ribosomal protection; partial reference coverage, therefore interpreted cautiously.

gyrA S83L	Strict; protein variant model; Curated-R	99.89	100	Fluoroquinolone resistance-associated target alteration.
PBP3 D350N/S357N	Strict; *H. influenzae* PBP3 protein variant model; Curated-R	53.29	96.39	β-lactam resistance-associated variant predicted by RGI; requires independent validation against *E. coli ftsI/PBP3*.
AcrAB-TolC-associated resistance model	Strict; protein overexpression model	100	100	Efflux-associated resistance model involving AcrAB-TolC; interpreted as a resistance-associated efflux mechanism rather than an acquired AMR gene.
MarR-associated resistance model	Strict; protein overexpression model; Curated-R	97.92	100	Regulatory alteration associated with efflux-mediated resistance; model-based prediction.
SoxS-associated resistance model	Strict; protein overexpression model	100	100	Regulatory/efflux-associated resistance model; model-based prediction.

3.7. Antimicrobial susceptibility and genotype–phenotype relationship

Table 4: Antimicrobial susceptibility phenotype and associated genomic determinants of *E. coli* BCSIR SA 01

Antimicrobial	Disc (µg)	Phenotype	Zone diameter (mm)	Associated genomic determinant(s)	Genotype–phenotype interpretation
Ampicillin	10	Resistant	12	EC-5 β-lactamase; AcrAB-TolC-associated determinants; PBP3 D350N/S357N*	Genomic findings are consistent with β-lactam resistance; the individual contribution of each determinant was not

					experimentally established.
Ceftazidime	30	Resistant	0	EC-5 β-lactamase; AcrAB-TolC-associated determinants; PBP3 D350N/S357N*	Genomic findings are consistent with the observed resistant phenotype; causal contribution of individual determinants was not established.
Ceftriaxone	30	Resistant	17	EC-5 β-lactamase; AcrAB-TolC-associated determinants; PBP3 D350N/S357N*	Genomic findings are consistent with the observed resistant phenotype; causal contribution of individual determinants was not established.
Nalidixic acid	30	Resistant	12	**gyrA S83L**; AcrAB-TolC-associated determinants	The *gyrA* S83L variant provides a resistance-associated genomic explanation for the observed phenotype.
Amikacin	30	Susceptible	20	No acquired aminoglycoside resistance determinant reported	Susceptible phenotype; no acquired AMR determinant was identified in the reported analysis.
Gentamicin	10	Susceptible	23	No acquired aminoglycoside resistance determinant reported	Susceptible phenotype; no acquired AMR determinant was identified in the reported analysis.

***** The PBP3 D350N/S357N variant was identified by the CARD/RGI *Haemophilus influenzae* PBP3 variant model and showed 53.29% identity; therefore, its role in the *E. coli* phenotype requires independent validation against the *E. coli ftsI/PBP3* sequence.

3.8. Virulence-associated genomic determinants

Table 5. Major virulence-associated determinants identified in *E. coli* BCSIR SA 01

Virulence category	Major determinant(s)	Gene(s) identified	Genomic location
Exotoxins & cytotoxins	Alpha-hemolysin	*hlyA-hlyD*	NODE_53
	Cytotoxic necrotizing factor 1	*cnf1*	NODE_53
	Colibactin	*clbA, clbB-clbS*	NODE_70 / NODE_32
	Vacuolating autotransporter toxin	*vat*	NODE_23
	Pic serine protease	*pic*	NODE_100
Adherence & fimbriae	P fimbriae	*papB, papC, papD, papE, papF, papH, papJ, papK*	NODE_85
	S/F1C fimbriae	*sfaD-sfaH, sfaS, sfaY*	NODE_86
	Type 1 fimbriae	*fimA-fimI*	NODE_24
	Common pilus / curli	*ecp*, *csgA-csgG*	NODE_23 / NODE_3
Iron acquisition	Enterobactin	*ent/fep* genes	NODE_1
	Salmochelin	*iroBCDEN*	NODE_76
	Yersiniabactin	*ybt/fyuA/irp* genes	NODE_44
	Heme uptake	*chuA-chuY*	NODE_2
Immune evasion	Capsule	*kpsC, kpsD, kpsE, kpsF, kpsM, kpsS, kpsU*	NODE_29
Invasion	Invasion-associated determinants	*ibeB, ibeC, aslA*	NODE_1 / NODE_16 / NODE_13

Motility & chemotaxis	Flagellar/chemotaxis machinery	*mot, flh, flg, fli, che* genes	Multiple nodes
Secretion systems	Type II secretion system	*gspC-gspM*	NODE_29
	Type VI secretion system	*tssD2/hcp2, tssJ, tssL, clpV/tssH*	NODE_48

3.9. Genomic islands and pathogenicity-associated regions

Table 6 summarizes the major genomic islands and pathogenicity-associated regions identified within the draft genome of *E. coli* BCSIR SA 01 using IslandViewer 4, revealing a substantial accessory genome equipped for virulence, adhesion, and survival. The analysis identified several high-profile pathogenicity islands (PAIs), most notably a 36.0 kb PAI-like region (NODE_53) harboring the *hlyABCD* operon, which encodes a hemolysin secretion system, and *cnf1*, which encodes cytotoxic necrotizing factor 1. The genome also contains a 30.4 kb capsule-associated island carrying the *kps* gene cluster (*kpsC, D, E, S, T, M*), indicating a genetic basis for polysaccharide capsule biosynthesis and export. Furthermore, the strain possesses a large 70.3 kb island containing the *iroBCDEN* salmochelin gene cluster, which is associated with glucosylated enterobactin production and enhanced iron acquisition—a critical virulence trait. A notably large 148.5 kb genomic island was identified (NODE_85) that combines multiple virulence and resistance features, including P-fimbrial adhesin genes (*papI, C, E, F*), the tetracycline resistance determinant *tet(B)*, and components of a Type VI Secretion System (T6SS). Additional regions include an EspC-like autotransporter with an associated tetracycline resistance gene and a Colicin I receptor (*cirA*), which may aid in competitive interactions. Collectively, the presence of these mobile genetic elements and virulence-associated clusters strongly supports the pathogenic potential of *E. coli* BCSIR SA 01 and aligns with the *in silico* pathogenicity predictions.

Table 6: Selected genomic island regions identified in *Escherichia coli* BCSIR SA 01 by IslandViewer 4 and representative genes/features within the predicted regions.

Genomic island / region	**Coordinates (bp)***	**Size (kb)**	**Associated contig / alignment context**	**Representative genes / features identified**	**Interpretation / biological relevance**
PAI-like genomic island region	716,881–752,877	36.0	NODE_53; reference-aligned context	*hlyA, hlyB, hlyC, hlyD, cnf1, vasL, emrE, ISSod13*	Genomic region containing hemolysin- and CNF1-associated virulence

					determinants together with secretion/mobile-element features.
Capsule-associated genomic island region	3,178,403–3,208,827	30.4	NODE_29	*kpsC, kpsD, kpsE, kpsS, kpsT, kpsM*, IS3/IS911, *yeeV*	Region containing capsule biosynthesis/export-associated *kps* genes, supporting a capsule-associated phenotype.
Salmochelin-associated genomic island region	5,369,555–5,439,842	70.3	NODE_76; **iroBCDEN cluster focus: 5,404–5,420 kb**	*iroB, iroC, iroD, iroE, iroN, papX*, transposase	Large predicted region containing the *iroBCDEN* salmochelin gene cluster, consistent with glucosylated enterobactin production and iron-acquisition functions.
Large pap/tet(T6SS)-associated genomic island region	5,093,776–5,242,321	148.5	NODE_85; **focus: 5,144–5,195 kb**	*papI, papC, papE, papF, papX, tetR, tet(B)*, T6SS-associated *tagF, pppA*, IS911	Large predicted region encompassing P-fimbrial/adhesin-associated genes, the tetracycline resistance determinant *tet(B)*, T6SS-associated genes, and mobile elements.
EspC-containing autotransporter region	329,846–352,361	22.5	Draft scaffold	*espC*-like serine protease autotransporter, *tetA(P)*, O-	Predicted genomic region containing an EspC-like autotransporter and additional resistance/surface-

				antigen acetylase	associated features; functional expression was not assessed.
Colicin I receptor-containing region	2,113,173–2,122,764	9.6	Draft scaffold	**Colicin I receptor (*cirA*)**	Predicted region containing a Colicin I receptor, consistent with a bacteriocin-uptake/competitive interaction function.

Coordinates and genomic regions are reported as provided by the IslandViewer 4 analysis of the submitted draft genome. Reference-alignment context is indicated where applicable.

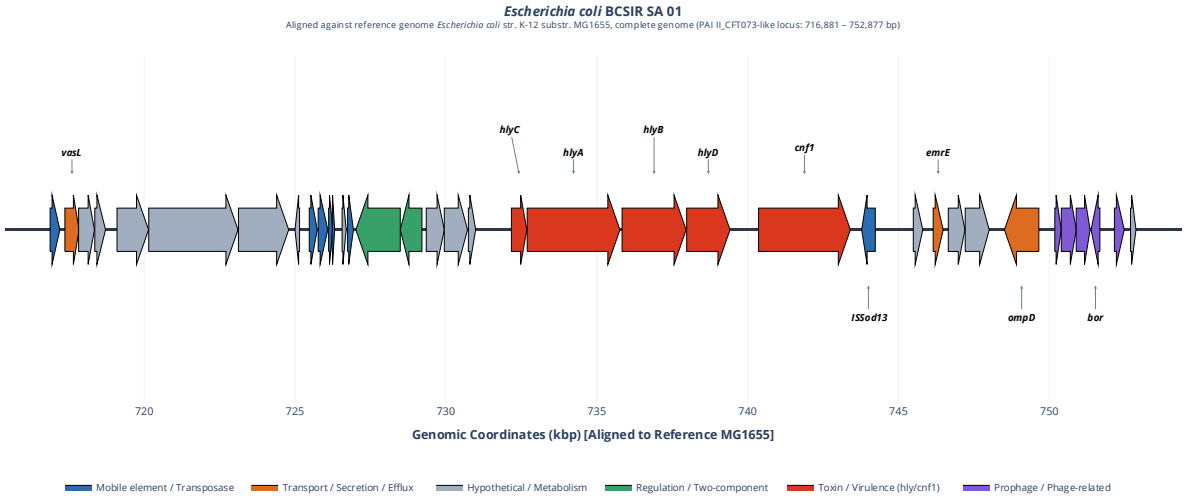


Figure 4: Genetic architecture of the 36.0-kb PAI-like pathogenicity island in *Escherichia coli* BCSIR SA 01

Genomic organization of the predicted 36.0-kb PAI-like region located at 716,881–752,877 bp, showing representative virulence-associated genes including *hlyA-hlyD* and *cnf1*, together with secretion-, efflux-, and mobile-element-associated features. The region is presented in the reference-aligned context reported by IslandViewer 4.

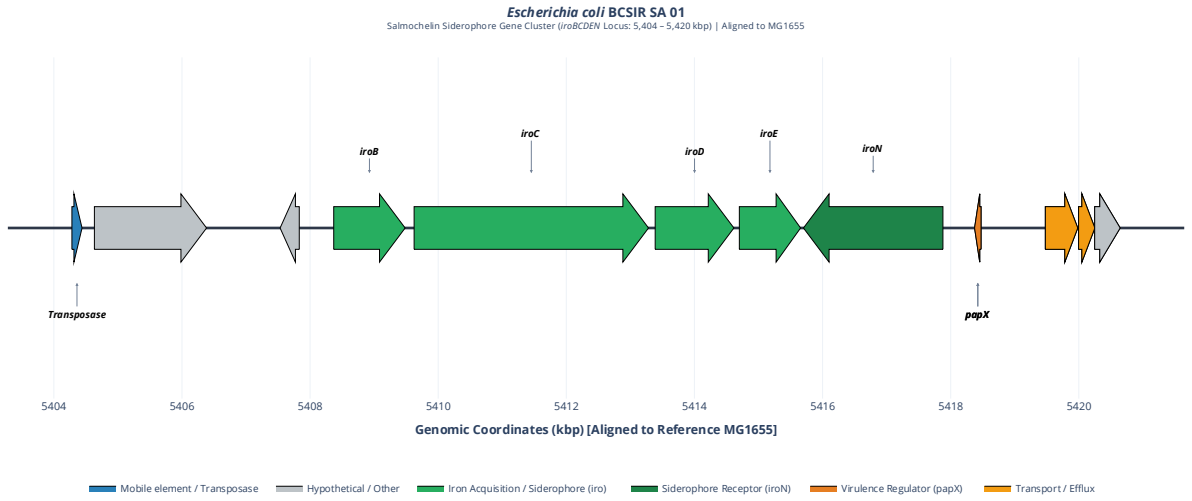


Figure 5: Genetic organization of the salmochelin siderophore cluster (*iroBCDEN*) in *Escherichia coli* BCSIR SA 01.

Genomic organization of the predicted salmochelin-associated region, highlighting the *iroB, iroC, iroD, iroE,* and *iroN* genes at the principal *iroBCDEN* cluster focus (approximately 5,404–5,420 kb), together with adjacent mobile-element and regulatory features. Coordinates are shown in the reference-aligned context provided by the analysis.

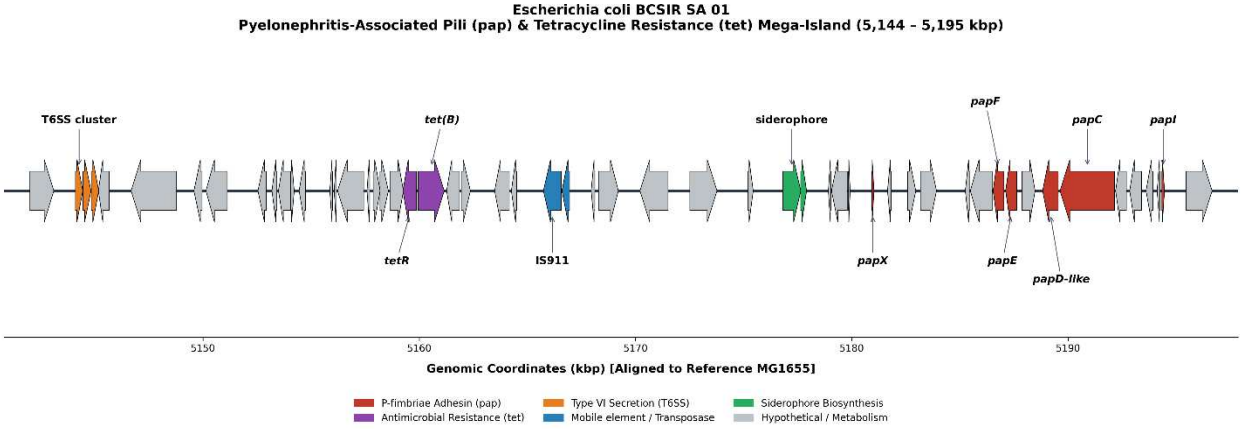


Figure 6: Genetic architecture of the multidrug resistance and virulence mega-island (*pap/tet(B)* locus) in *Escherichia coli* BCSIR SA 01.

Genomic organization of the predicted 148.5-kb pap/tet/T6SS-associated region, showing representative *pap* fimbrial/adhesion-associated genes, *tetR* and *tet(B)* resistance determinants, T6SS-associated genes (*tagF, pppA*), and the mobile element IS911. The region spans 5,093,776–5,242,321 bp, with a principal focus at approximately 5,144–5,195 kb.

3.10. Mobile genetic elements and phage-associated features

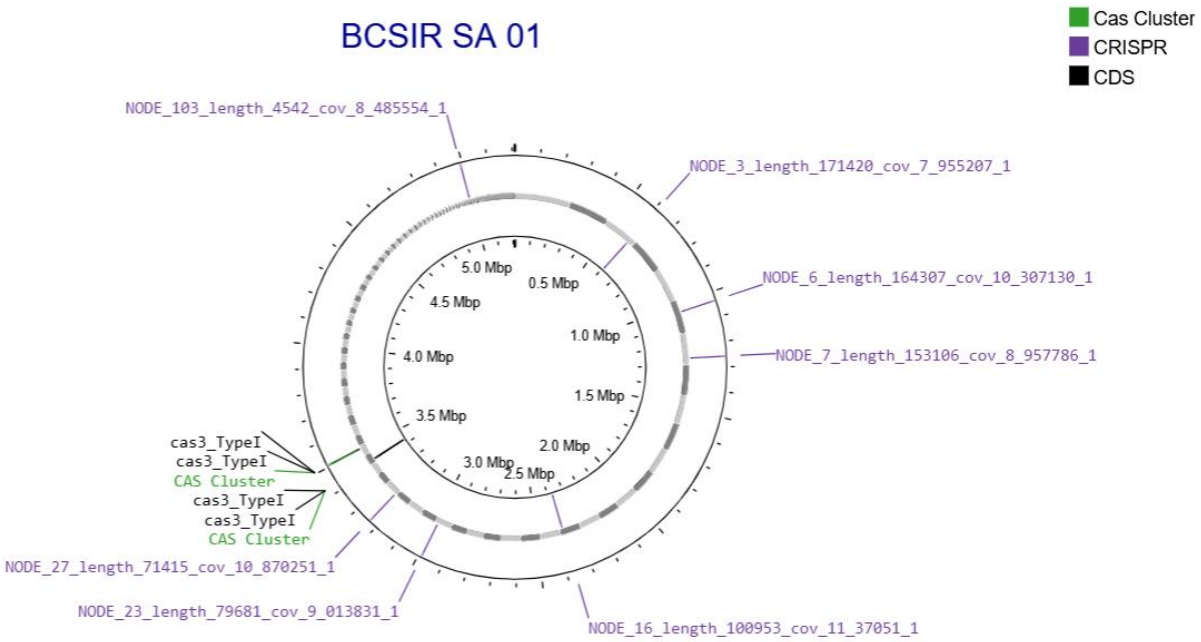


Figure 7A: CRISPR-Cas analysis features identified in *Escherichia coli* BCSIR SA 01.

CRISPR-Cas-associated genomic regions identified in *E. coli* BCSIR SA 01, showing the locations of predicted CRISPR and Cas-associated features across the draft genome

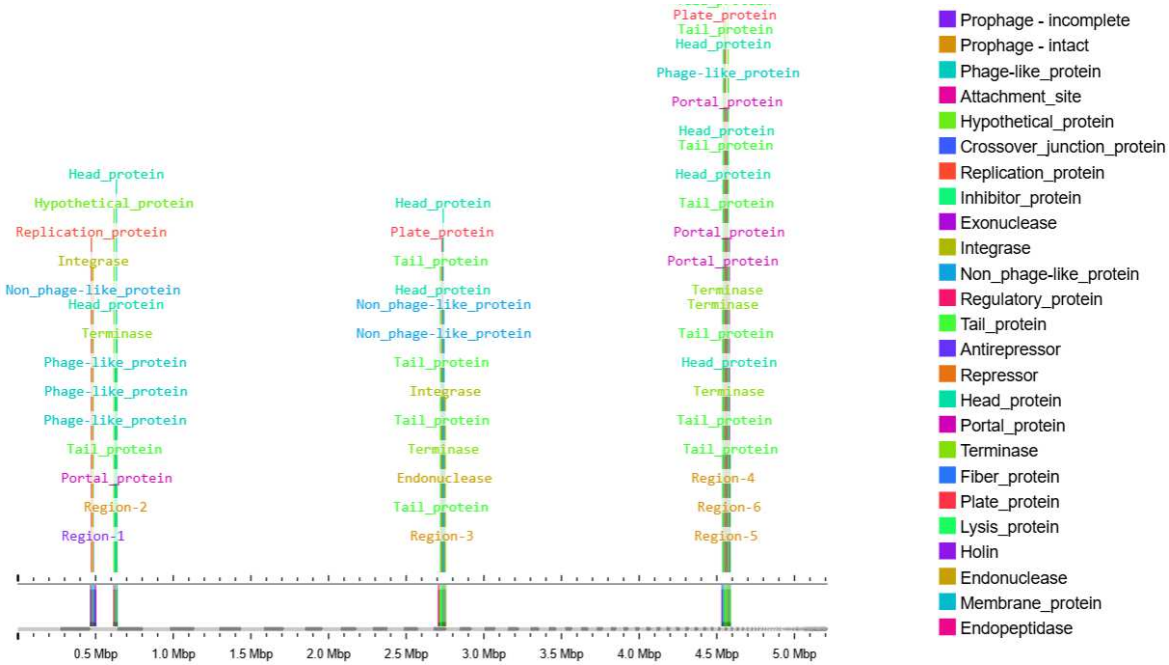


Figure 7B: PHASTEST analysis features identified in *Escherichia coli* BCSIR SA 01.

PHASTEST-based identification and genomic localization of predicted prophage- and phage-associated regions in *E. coli* BCSIR SA 01.

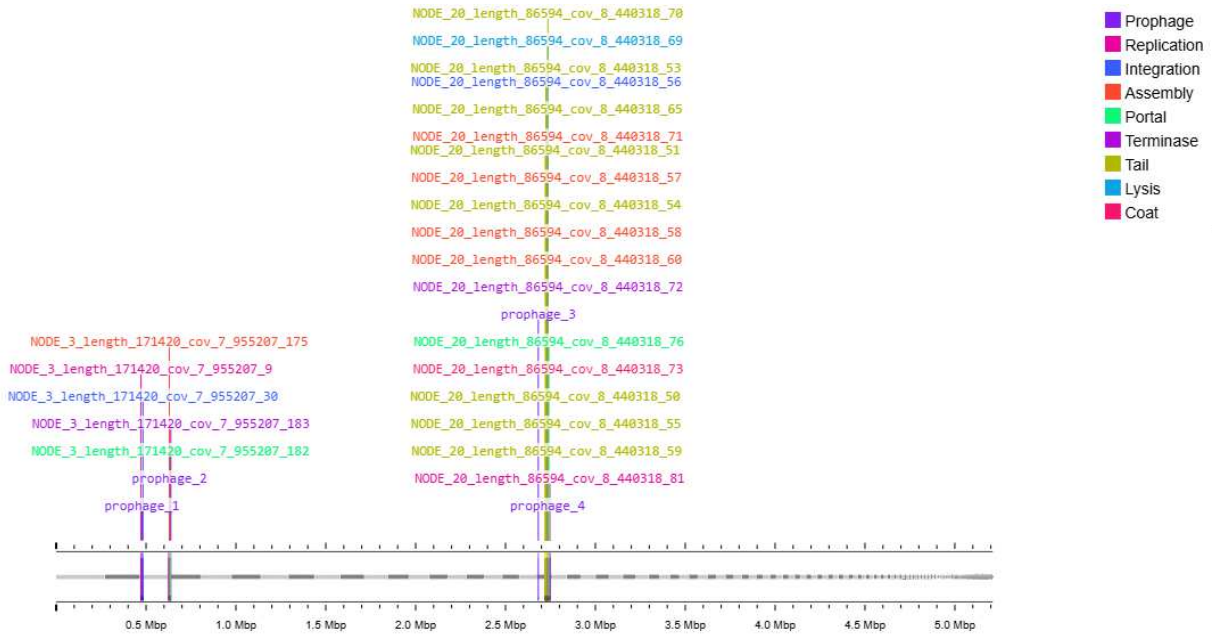


Figure 7C: PHIGARO analysis features identified in *Escherichia coli* BCSIR SA 01.

PHIGARO-based identification of predicted phage-associated genomic regions, showing the distribution of phage-related functional features across the *E. coli* BCSIR SA 01 genome.

Distribution of Viral and Phage-Associated Elements

Figure 7D illustrates the distribution of viral and phage-associated genomic sequences within the draft genome of *E. coli* BCSIR SA 01, as predicted by the VirSorter tool. The circular plot maps the genomic architecture, displaying GC content and GC skew alongside the identified viral features. VirSorter analysis detected multiple viral or prophage-associated regions (indicated in green), which are distributed across various contigs of varying lengths. Notably, these predicted regions are present on both large contigs, such as NODE_3 (171 kb) and NODE_62/61 (~20-21 kb), as well as numerous smaller contigs. The widespread distribution of these viral signatures across the draft genome suggests a history of multiple phage integration events. The presence of these integrated viral elements (prophages) is biologically significant, as they often carry accessory genes, including virulence factors or antibiotic resistance determinants, that can contribute to the pathogenicity, adaptability, and evolutionary trajectory of the bacterial strain.

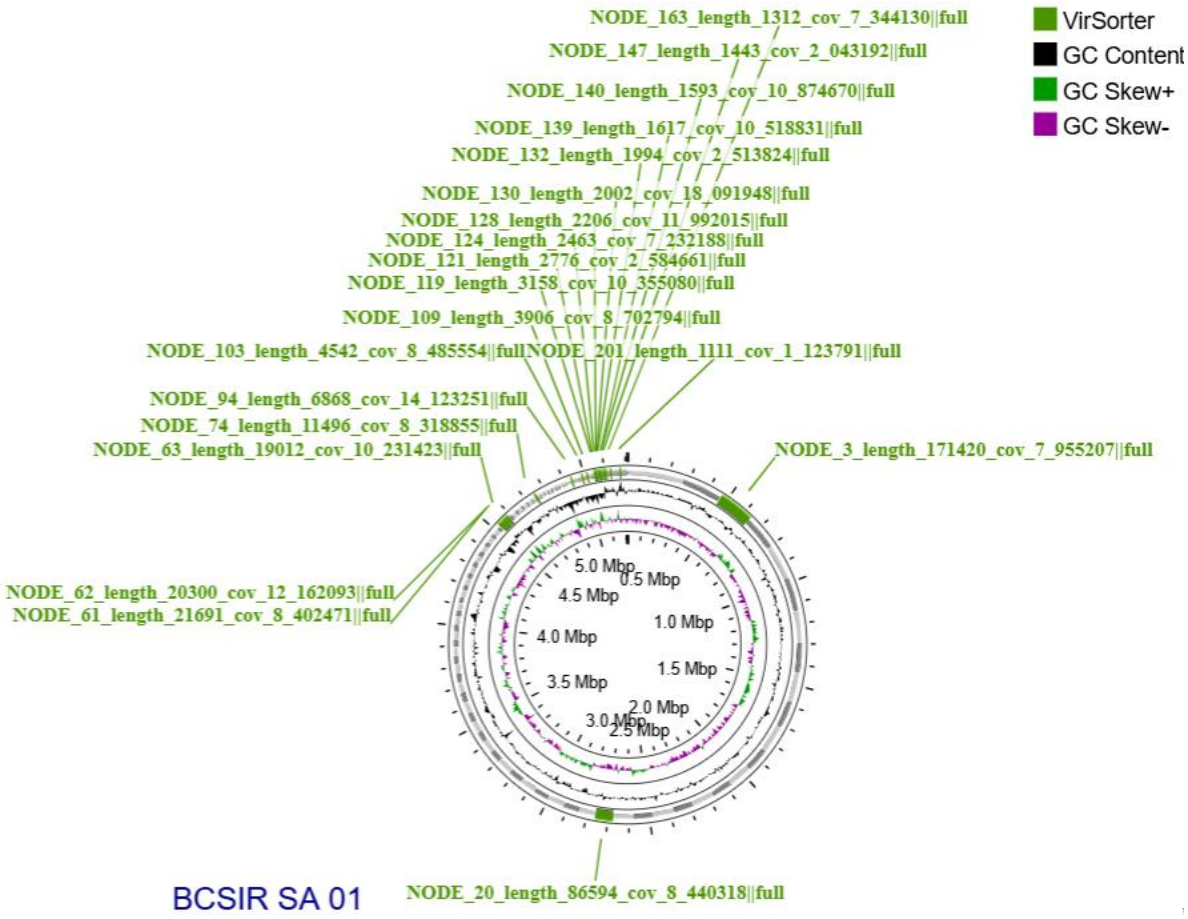
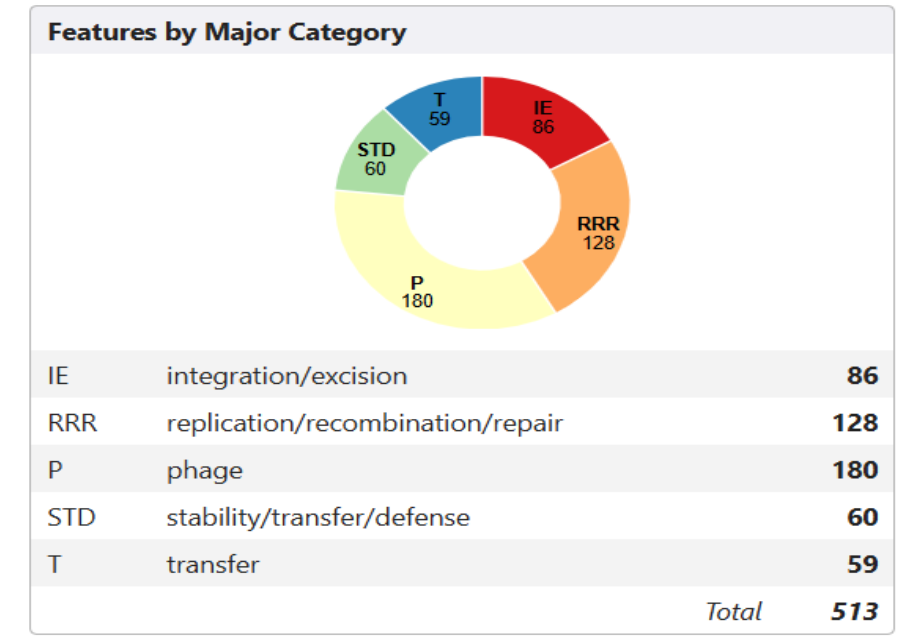


Figure 7D: VirSorter-based prediction of viral and phage-associated genomic sequences in *E. coli* BCSIR SA 01. The circular genome map displays the distribution of predicted viral sequences across the draft genome. The outer rings represent the contigs, while the inner track indicates GC content (black) and GC skew (green/purple). VirSorter-predicted prophage or viral regions are highlighted in green and mapped to their respective contigs (NODE).

Figure 7E presents the functional categorization of mobile genetic element (MGE)-associated features identified within the genome of *E. coli* BCSIR SA 01, classified using a database-driven approach (likely MobileElementFinder or a similar tool). A total of 513 MGE-related features were detected and grouped into five major functional categories. The largest proportion of these features is associated with phage functions (n = 180), followed closely by genes involved in replication, recombination, and repair (RRR; n = 128). Additionally, the genome contains a substantial number of features related to integration/excision (IE; n = 86) and stability, transfer, and defense (STD; n = 60), as well as those directly involved in conjugative transfer (T; n = 59). This diverse functional repertoire indicates that the genome of *E. coli* BCSIR SA 01 is highly enriched with mobile genetic elements. The abundance of phage and recombination-associated features suggests a high degree of genomic plasticity, likely facilitating horizontal gene transfer, genomic rearrangement, and the potential acquisition of advantageous traits such as virulence factors or antimicrobial resistance genes.



IE	integration/excision	**86**
RRR	replication/recombination/repair	**128**
P	phage	**180**
STD	stability/transfer/defense	**60**
T	transfer	**59**
	Total	***513***

Figure 7E: Functional categorization of mobile genetic element (MGE) features identified in *E. coli* BCSIR SA 01. The donut chart and accompanying table summarize the distribution of 513 total features associated with mobile genetic elements, categorized by major functional roles: phage (P), replication/recombination/repair (RRR), integration/excision (IE), stability/transfer/defense (STD), and transfer (T).

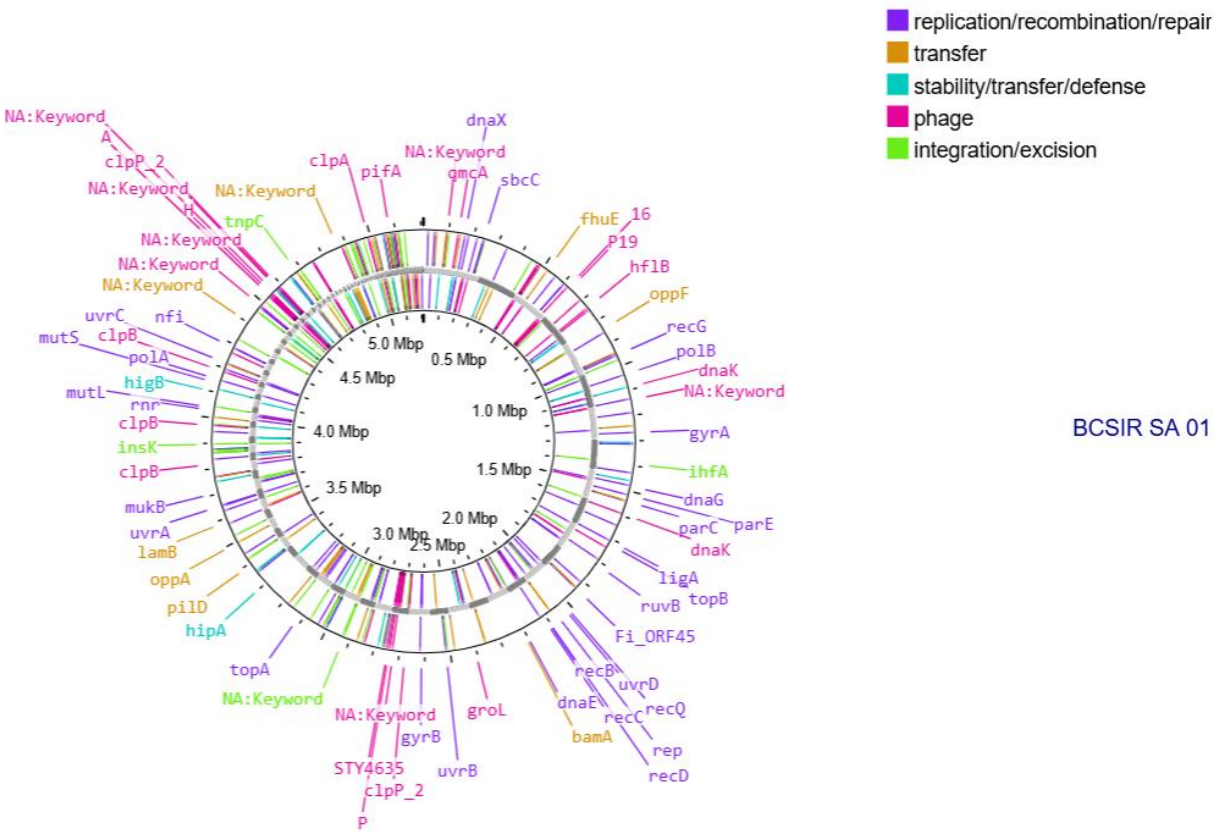


Figure 7F: Distribution of mobile genetic element-associated features identified by mobileOG-db in *E. coli* BCSIR SA 01. The circular genome map illustrates the chromosomal locations of 513 mobile genetic element (MGE)-associated features. Features are color-coded by major functional categories: replication/recombination/repair (purple), phage (pink/magenta), integration/excision (light green), stability/transfer/defense (cyan), and transfer (orange). Representative genes and features from each category are annotated.

Figure 7F illustrates the chromosomal distribution and functional diversity of mobile genetic element (MGE)-associated features in E. coli BCSIR SA 01, as characterized by the mobileOG-db database. A total of 513 MGE-related features were mapped across the draft genome and categorized into five major functional groups: replication/recombination/repair (n = 128), phage (n = 180), integration/excision (n = 86), stability/transfer/defense (n = 60), and transfer (n = 59). The genome map reveals that these features are distributed broadly across the entire chromosome rather than being confined to a single locus. The annotations highlight a diverse array of genes essential for genomic plasticity and adaptation. For instance, the replication and repair category includes key genes such as polA, mutL, uvrD, recQ, and dnaK, which are crucial for maintaining genome integrity and facilitating homologous recombination. The phage and integration/excision categories include numerous integrases, transposases, and phage-associated genes (e.g., STY4635, clpP), indicating historical and ongoing phage infections. Furthermore, the presence of genes associated with transfer and stability (e.g., pilD, bamA) supports the strain's capacity for horizontal

gene transfer. Collectively, this widespread distribution of MGEs underscores the genomic fluidity of *E. coli* BCSIR SA 01 and its potential to acquire or disseminate clinically relevant traits such as virulence factors or antibiotic resistance determinants.

3.11. Functional genome annotation

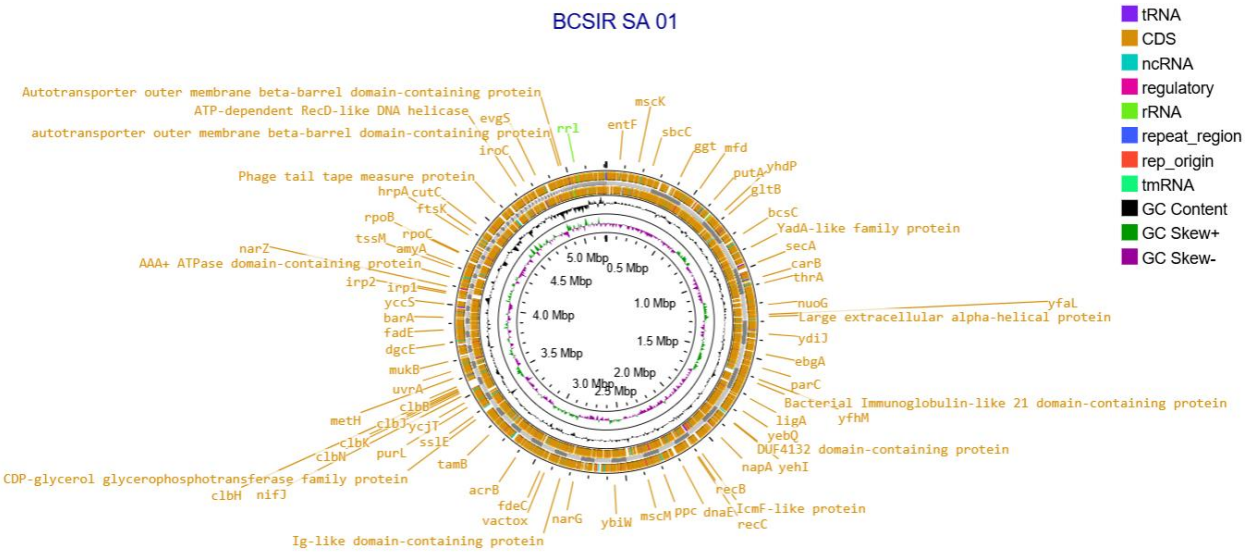


Figure 8A: Circular genome map of *E. coli* BCSIR SA 01. The outermost rings display annotated functional genes and coding sequences (CDS) on both strands, followed by RNA and regulatory features, repeat regions, and the replication origin. The innermost tracks depict GC content (black) and GC skew (green/purple) across the genome.

The circular genome map of *E. coli* BCSIR SA 01 illustrates the distribution of functional genes, coding sequences, and regulatory elements across the ~5.21 Mb draft genome. The outer rings highlight key annotated genes, revealing a diverse repertoire of virulence-associated, metabolic, and membrane-associated proteins. The inner tracks show GC content and GC skew, with notable fluctuations that often correspond to horizontally acquired regions. Overall, the genome visualization demonstrates a densely packed and structurally organized chromosome harboring numerous accessory features (Figure 8A).

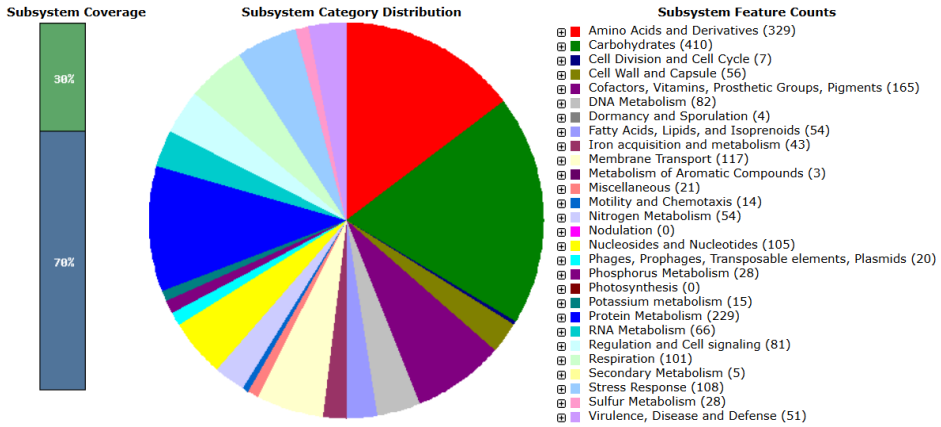


Figure 8B. RAST-based functional annotation of *Escherichia coli* BCSIR SA 01. The pie chart illustrates the distribution of genes across various subsystem categories, while the accompanying bar indicates the overall subsystem coverage. The legend on the right provides the specific feature counts for each functional category.

The functional annotation of the *E. coli* BCSIR SA 01 genome was performed using the RAST server, resulting in a subsystem coverage of approximately 70%. The distribution of annotated genes reveals that the largest proportions are assigned to "Amino Acids and Derivatives" (329), "Protein Metabolism" (229), and "Carbohydrates" (410). Notably, functional categories related to pathogenicity and adaptation are also present, including "Virulence, Disease and Defense" (51), "Membrane Transport" (117), and "Iron acquisition and metabolism" (43), which align with the strain's predicted pathogenic nature.

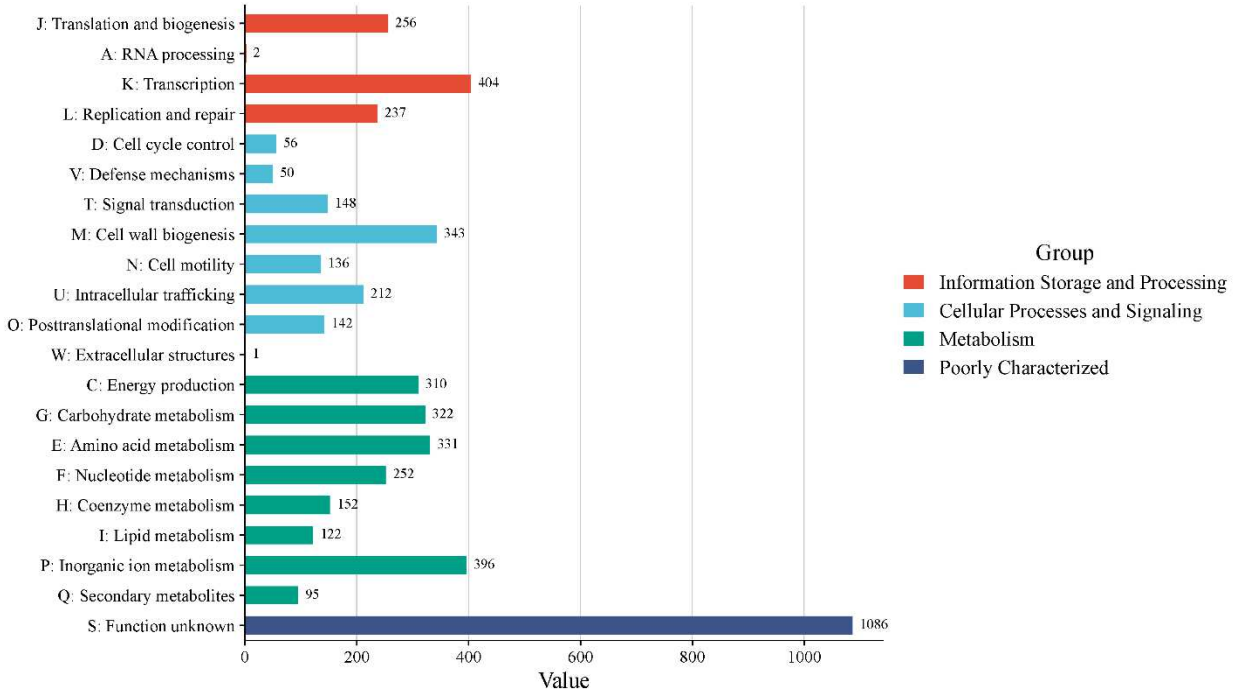


Figure 8C. COG-based functional classification of *Escherichia coli* BCSIR SA 01. The bar chart displays the distribution of genes assigned to Clusters of Orthologous Genes (COG) categories. The x-axis represents the number of genes (value), and the y-axis lists the specific COG functional categories. The bars are color-coded into four major functional groups: Information Storage and Processing (red), Cellular Processes and Signaling (light blue), Metabolism (teal), and Poorly Characterized (dark blue).

The functional classification of the E. coli BCSIR SA 01 genome was performed using the Clusters of Orthologous Genes (COG) database, as shown in Figure 8C. A large proportion of genes were assigned to essential metabolic functions (teal), with the highest counts observed in "Inorganic ion metabolism" (396), "Cell wall biogenesis" (343), "Amino acid metabolism" (331), and "Carbohydrate metabolism" (322). Within the Information Storage and Processing group (red), genes related to "Transcription" (404) and "Translation and biogenesis" (256) were highly represented. Notably, genes associated with "Defense mechanisms" (50) and "Cell motility" (136) were also identified. Furthermore, a substantial number of genes (1,086) fell into the "Function unknown" category, which is typical for newly sequenced bacterial genomes and highlights areas for future functional characterization.

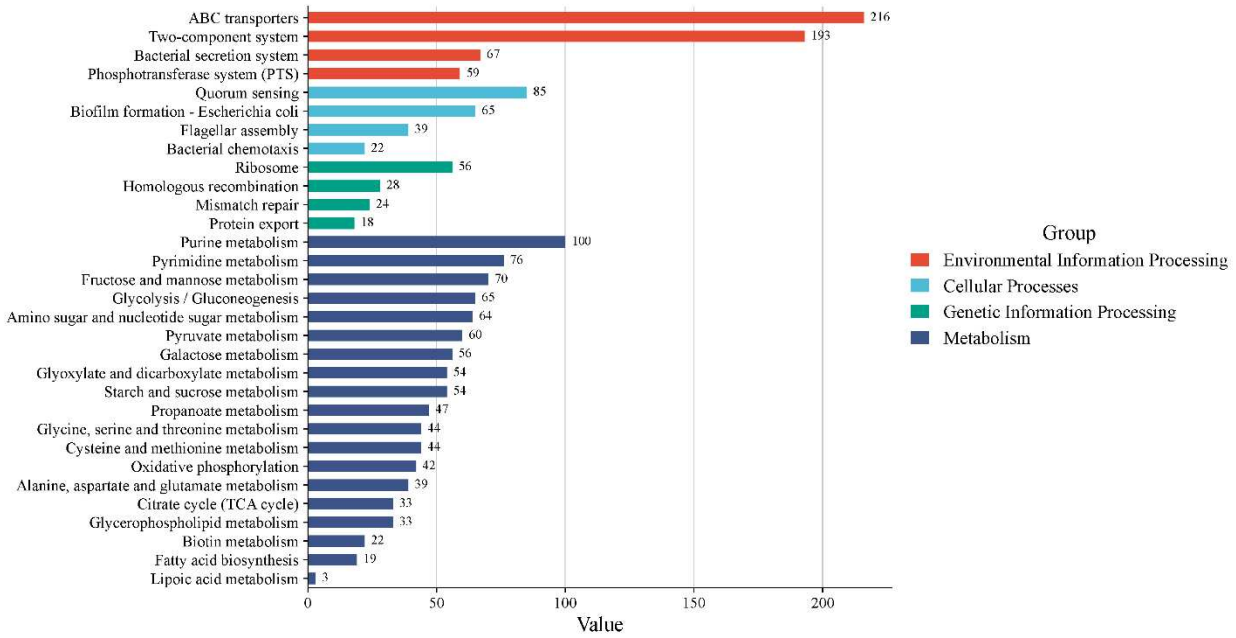
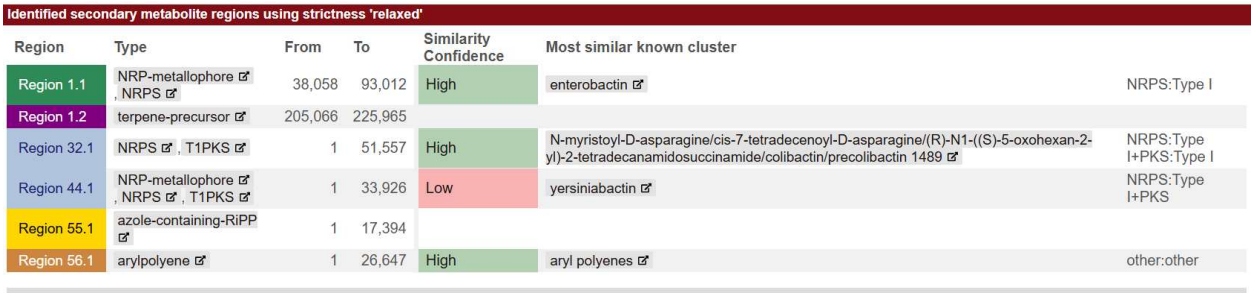


Figure 8D. KEGG-based functional annotation of Escherichia coli BCSIR SA 01. The horizontal bar chart illustrates the distribution of genes mapped to specific Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The x-axis represents the number of genes, and the y-axis lists the pathway categories. The bars are color-coded into four functional groups: Environmental Information Processing (red), Cellular Processes (light blue), Genetic Information Processing (teal), and Metabolism (dark blue). KEGG pathway database entries are reproduced and annotated with permission from Kanehisa Laboratories.

The functional potential of the *E. coli* BCSIR SA 01 genome was further analyzed using the KEGG pathway database (Figure 8D). The majority of the annotated genes were assigned to metabolic pathways (dark blue), with prominent representation in "Purine metabolism" (100), "Pyrimidine metabolism" (76), and various carbohydrate and amino acid metabolism pathways. Within the Environmental Information Processing category (red), "ABC transporters" (216) and "Two-component systems" (193) were the most abundant, reflecting the strain's robust capacity for nutrient uptake and environmental sensing. Furthermore, genes associated with "Bacterial secretion systems" (67) and "Biofilm formation" (65) were identified, which are critical mechanisms for host interaction and virulence. The presence of these specific pathways supports the genomic evidence that *E. coli* BCSIR SA 01 is well-equipped for survival, adaptation, and pathogenic behavior.

3.12. Secondary metabolite biosynthetic potential



Identified secondary metabolite regions using strictness 'relaxed'

Region	Type	From	To	Similarity Confidence	Most similar known cluster	
Region 1.1	NRP-metallophore, NRPS	38,058	93,012	High	enterobactin	NRPS:Type I
Region 1.2	terpene-precursor	205,066	225,965			
Region 32.1	NRPS, T1PKS	1	51,557	High	N-myristoyl-D-asparagine/cis-7-tetradecenoyl-D-asparagine/(R)-N1-((S)-5-oxohexan-2-yl)-2-tetradecanamidosuccinamide/colibactin/precolibactin 1489	NRPS:Type I+PKS:Type I
Region 44.1	NRP-metallophore, NRPS, T1PKS	1	33,926	Low	yersiniabactin	NRPS:Type I+PKS
Region 55.1	azole-containing-RiPP	1	17,394			
Region 56.1	arylpolyene	1	26,647	High	aryl polyenes	other:other

Figure 9: Predicted secondary metabolite biosynthetic gene clusters identified by antiSMASH in *E. coli* BCSIR SA 01.

antiSMASH analysis identified multiple predicted secondary metabolite biosynthetic regions in *E. coli* BCSIR SA 01, with varying levels of similarity and confidence to known biosynthetic gene clusters. These findings indicate the presence of genomic regions potentially involved in secondary metabolite biosynthesis.

4. Discussion

The present study provides an integrated phenotypic and whole-genome characterization of *Escherichia coli* BCSIR SA 01, revealing a combination of antimicrobial resistance-associated determinants, multiple virulence-associated genes, and a diverse accessory-genome profile. The draft genome was 5.21 Mb in size and contained 234 contigs, with 100% estimated completeness and 1.39% contamination, supporting its suitability for downstream genomic analyses despite the fragmented assembly. (Gurevich et al., 2013; Chklovski et al., 2023). The isolate was assigned to Clermont phylogroup B2, serotype O6:H31, MLST ST127 and *fimH* type fimH180, with 98.61% FastANI similarity supporting its species-level classification as *E. coli*. The ST127/O6:H31/B2 combination is particularly relevant because ST127 has been repeatedly described among uropathogenic *E. coli* and has been associated with diverse virulence and antimicrobial resistance determinants. (Clermont et al., 2013; Joensen et al., 2015; Cave et al., 2022). Phenotypically, BCSIR SA 01 was resistant to ampicillin, ceftazidime, ceftriaxone and nalidixic acid, while remaining susceptible to amikacin and gentamicin. CARD/RGI analysis identified an EC-5 β-lactamase homolog with high sequence identity and complete reference coverage, providing a plausible genomic explanation for the observed β-lactam resistance. (Pitout, 2012; Alcock *et al*., 2023). The *gyrA* S83L substitution provides an additional resistance-associated explanation for the nalidixic-acid-resistant phenotype, as alterations in *gyrA* are well-established contributors to quinolone resistance in *E. coli*. (Yao et al., 2007). The presence of *tet(B)* further indicates tetracycline-resistance potential, although tetracycline susceptibility was not assessed phenotypically in this study. (Nelson and Levy, 1999). The genome also contained AcrAB-TolC-, MarR- and SoxS-associated resistance models, indicating the possible contribution of efflux and regulatory mechanisms to the isolate's resistance profile. However, these computational predictions should not be interpreted in the same manner as experimentally confirmed acquired

resistance genes. (McArthur et al., 2013; Alcock et al., 2023). Similarly, the PBP3 D350N/S357N prediction requires caution because it was detected using a *Haemophilus influenzae* PBP3 model and showed only 53.29% identity, and therefore its contribution to resistance in *E. coli* remains uncertain. (Alcock et al., 2023). A broad virulence-associated repertoire was identified, including *hlyA-hlyD*, *cnf1*, *pap*, *sfa*, *fim*, *iroBCDEN*, *ybt/fyuA/irp*, *chu*, *kps*, and multiple secretion and motility-associated genes. Such combinations are consistent with the multifactorial nature of UPEC pathogenicity, in which adhesion, iron acquisition, immune evasion, toxin production and host adaptation are mediated by multiple interacting determinants rather than a single virulence factor. (Wiles, Kulesus and Mulvey, 2008; Subashchandrabose and Mobley, 2015). The detection of *iroBCDEN* is particularly notable because salmochelin-mediated iron acquisition is an established feature of ExPEC and can provide an advantage under host iron limitation. (Fischbach et al., 2006; Russo et al., 2008).

IslandViewer analysis further revealed several predicted genomic regions containing virulence-, resistance-, capsule-, siderophore- and mobile-element-associated genes. The 36.0-kb PAI-like region containing *hlyA-hlyD* and *cnf1* and the 148.5-kb pap/tet-associated region containing *pap* genes, *tet(B)* and T6SS-associated features were particularly notable. Genomic islands are important contributors to bacterial genome plasticity and may carry clusters of horizontally acquired genes; however, because the present genome is a draft assembly, the predicted regions should not be regarded as experimentally confirmed physically linked structures. (Hacker and Kaper, 2000; Bertelli et al., 2017).

The additional detection of prophage-, viral- and mobile-element-associated features suggests considerable accessory-genome diversity in BCSIR SA 01. Such elements can contribute to bacterial genome plasticity and may facilitate acquisition or rearrangement of resistance and virulence-associated determinants. (Arndt *et al*., 2019; Brown *et al*., 2022). The CRISPR-associated features identified in the genome further indicate the presence of a defense-related genomic system, although their activity and functional contribution were not experimentally assessed. (Couvin et al., 2018).

The overall genomic profile of BCSIR SA 01 is consistent with previous studies demonstrating that WGS can reveal resistance determinants, virulence factors, typing characteristics and accessory-genome features simultaneously. (Vanstokstraeten et al., 2023). Importantly, the present findings also demonstrate why genomic and phenotypic data should be interpreted together, because detection of a resistance-associated gene or mutation does not necessarily establish its direct contribution to the observed phenotype. (Vanstokstraeten et al., 2023).

In the context of Bangladesh, where resistant and ESBL-producing *E. coli* have been documented in clinical and healthcare-associated settings, the genomic characterization of individual isolates can contribute to a better understanding of clinically relevant resistance and virulence profiles. (Moniruzzaman et al., 2023). However, because this study investigated a single isolate, the findings should not be generalized to the wider UPEC population, and larger comparative genomic

datasets will be required to determine the epidemiological distribution and broader significance of the observed ST127-associated genomic profile.

5. Conclusion

This study provided an integrated phenotypic and whole-genome characterization of *Escherichia coli* BCSIR SA 01 isolated from a clinical urinary tract infection case. The isolate belonged to the B2 phylogroup, serotype O6:H31, MLST ST127, and FimH type fimH180, with 98.61% FastANI supporting its species-level identification as *E. coli*.

Genomic analysis identified multiple resistance-associated determinants, including an EC-5 β-lactamase homolog, *tet(B)*, *gyrA* S83L, and additional efflux- and regulatory-associated resistance models, which complemented the observed resistance to ampicillin, ceftazidime, ceftriaxone, and nalidixic acid. The genome also carried a broad repertoire of virulence-associated determinants involving toxins, adhesins, iron acquisition, capsule-associated functions, motility, invasion, and secretion systems. Several predicted genomic islands and mobile-element-associated regions contained combinations of virulence, resistance, and accessory-genome features, highlighting the genetic complexity of the isolate.

Overall, the findings demonstrate the value of combining antimicrobial susceptibility testing with whole-genome analysis to obtain a comprehensive profile of a clinically relevant UPEC isolate.

However, as this study was based on a single draft genome, the identified genomic features represent strain-level observations and should not be generalized to the wider UPEC population without analysis of larger comparative datasets.

References

1. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
2. Wiles, T. J., Kulesus, R. R. & Mulvey, M. A. Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Exp. Mol. Pathol.* **85**, 11–19 (2008).
3. Subashchandrabose, S. & Mobley, H. L. T. Virulence and fitness determinants of uropathogenic *Escherichia coli*. *Microbiol. Spectr.* **3**, UTI-0015-2012 (2015).
4. Johnson, J. R. Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* **4**, 80–128 (1991).
5. Chowdhury, S. F., Sarkar, M. M. H., Al Sium, S. M. et al. Metatranscriptomic insights into host microbiome interactions underlying asymptomatic COVID-19 cases. *Sci. Rep.* (2026). https://doi.org/10.1038/s41598-026-40563-x

6. Hacker, J. & Kaper, J. B. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679 (2000).

7. Pitout, J. D. D. Extraintestinal pathogenic *Escherichia coli*: an update on antimicrobial resistance, laboratory diagnosis and treatment. *Expert Rev. Anti-Infect. Ther.* **10**, 1165–1176 (2012).

8. World Health Organization. *Antimicrobial resistance*. (WHO, 2026).

9. McArthur, A. G., Waglechner, N., Nizam, F. et al. The Comprehensive Antibiotic Resistance Database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).

10. Cepas, V. & Soto, S. M. Relationship between virulence and resistance among Gram-negative bacteria. *Antibiotics* **9**, 719 (2020).

11. Islam, M. R., Hoque, M. J., Uddin, M. N., Dewan, A., Haque, N. B., Islam, M. T. et al. Antimicrobial resistance of *E. coli* causing urinary tract infection in Bangladesh. *Mymensingh Med. J.* **31**, 180–185 (2022).

12. Chowdhury, S. S., Tahsin, P., Xu, Y., Mosaddek, A. S. M., Muhamadali, H. & Goodacre, R. Trends in antimicrobial resistance of uropathogens isolated from urinary tract infections in a tertiary care hospital in Dhaka, Bangladesh. *Antibiotics* **13**, 925 (2024).

13. Moniruzzaman, M., Hussain, M. T., Ali, S., Hossain, M., Hossain, M. S., Alam, M. A. U. et al. Multidrug-resistant *Escherichia coli* isolated from patients and surrounding hospital environments in Bangladesh: a molecular approach for the determination of pathogenicity and resistance. *Heliyon* **9**, e22109 (2023).

14. Amin, M. B., Flatgard, B. M., Hoque, K. I., Khan, R., Islam, M. R., Rahman, M. et al. Foodborne sources of antimicrobial-resistant *Escherichia coli* in community-acquired urinary tract infections in Bangladesh. *Commun. Med.* **6** (2026).

15. Mohammad Mahmud, A. S., Andersson, P., Bulach, D., Duchene, S., daSilva, A. G., Lin, C. et al. Molecular epidemiology of SARS-CoV-2 in Bangladesh. *Viruses* **17**, 517 (2025).

16. Akter, S., Oliveira, J. I. N., Barton, C. et al. Spike protein mutations and structural insights of pangolin lineage B.1.1.25 with implications for viral pathogenicity and ACE2 binding affinity. *Sci. Rep.* **13**, 13146 (2023).

17. Vanstokstraeten, R., Piérard, D., Crombé, F. et al. Genotypic resistance determined by whole genome sequencing versus phenotypic resistance in 234 *Escherichia coli* isolates. *Sci. Rep.* **13**, 449 (2023).

18. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 124 (2018).

19. Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M. & Scheutz, F. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* **53**, 2410–2426 (2015).

20. Roer, L., Tchesnokova, V., Allesøe, R., Muradova, M., Chattopadhyay, S., Ahrenfeldt, J. et al. Development of a web tool for *Escherichia coli* subtyping based on *fimH* alleles. *J. Clin. Microbiol.* **55**, 2538–2543 (2017).

21. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

22. Cave, R., Ter-Stepanyan, M. M., Kotsinyan, N. & Mkrtchyan, H. V. An emerging lineage of uropathogenic extended spectrum β-lactamase *Escherichia coli* ST127. *Microbiol. Spectr.* **10**, e02511-22 (2022).

23. Johnson, J. R., Johnston, B., Clabots, C. R., Kuskowski, M. A., Roberts, E. & DebRoy, C. Virulence genotypes and phylogenetic background of *Escherichia coli* serogroup O6 isolates from humans, dogs, and cats. *J. Clin. Microbiol.* **46**, 417–422 (2008).

24. Leber, A. L. (ed.) *Clinical Microbiology Procedures Handbook*. 4th edn. (ASM Press, 2016).

25. Tajbakhsh, E., Tajbakhsh, S. & Khamesipour, F. Isolation and molecular detection of Gram negative bacteria causing urinary tract infection in patients referred to Shahrekord hospitals, Iran. *Iran. Red Crescent Med. J.* **17**, e24779 (2015).

26. Rahman, M. M., Hossain, M. M. K., Rubaya, R., Halder, J., Karim, M. E., Bhuiya, A. A. et al. Association of antibiotic resistance traits in uropathogenic *Escherichia coli* (UPEC) isolates. *Can. J. Infect. Dis. Med. Microbiol.* **2022**, 4251486 (2022).

27. Clinical and Laboratory Standards Institute (CLSI). *M02: Performance Standards for Antimicrobial Disk Susceptibility Tests*. 14th edn. (CLSI, 2024).

28. QIAGEN. *QIAamp DNA Mini and Blood Mini Handbook*. (QIAGEN, 2020).

29. Thermo Fisher Scientific. *Qubit 4 Fluorometer and dsDNA High-Sensitivity Assay Documentation*. (Thermo Fisher Scientific, 2019).

30. Thermo Fisher Scientific. *NanoDrop 2000/2000c Spectrophotometers User Manual*. (Thermo Fisher Scientific, 2009).

31. Illumina. *Illumina DNA Prep Reference Guide*. (Illumina, Inc., 2020).

32. Illumina. *MiniSeq System Guide*. (Illumina, Inc., 2021).

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

34. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).

35. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

36. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality. *Nat. Methods* **20**, 1203–1212 (2023).

37. Al Sium, S. M., Goswami, B., Chowdhury, S. F., Naser, S. R., Sarkar, M. K., Faruq, M. J. et al. An insight into the genome-wide analysis of bacterial defense mechanisms in a uropathogenic *Morganella morganii* isolate from Bangladesh. *PLoS ONE* **20**, e0313141 (2025).

38. Schwengers, O., Jelonek, L., Dieckmann, M. A., Beyvers, S., Blom, J. & Goesmann, A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genom.* **7**, 000685 (2021).

39. Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J. et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365 (2015).

40. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).

41. Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J. et al. Introducing the BV-BRC: a resource for bacterial and viral bioinformatics. *Nucleic Acids Res.* **51**, D678–D689 (2023).

42. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

43. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* 53, D672–D677 (2025).

44. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30 (2000).

45. Clermont, O., Christenson, J. K., Denamur, E. & Gordon, D. M. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **5**, 58–65 (2013).

46. Weissman, S. J., Johnson, J. R., Tchesnokova, V., Billig, M., Dykhuizen, D., Riddell, K. et al. High-resolution two-locus clonal typing of extraintestinal pathogenic *Escherichia coli*. *Appl. Environ. Microbiol.* **78**, 1353–1360 (2012).

47. Roer, L., Johannesen, T. B., Hansen, F., Stegger, M., Tchesnokova, V., Sokurenko, E. et al. CHTyper, a web tool for subtyping of extraintestinal pathogenic *Escherichia coli* based on the *fumC* and *fimH* alleles. *J. Clin. Microbiol.* **56**, e00063-18 (2018).

48. Alcock, B. P., Huynh, W., Chalil, R. et al. CARD 2023: expanded curation, support for machine learning and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699 (2023).

49. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917 (2022).

50. Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L. & Brinkman, F. S. L. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **45**, W30–W35 (2017).

51. Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D. S. PHAST, PHASTER and PHASTEST: tools for finding prophage in bacterial genomes. *Brief. Bioinform.* **20**, 1560–1567 (2019).

52. Starikova, E. V., Tikhonov, A. V., Prianichnikov, N. A. et al. Phigaro: high performance bacteriophage sequence identification in microbial genomes. *Bioinformatics* **36**, 3882–3884 (2020).

53. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).

54. Brown, C. L., Mullet, J., Hindi, F., Stoll, J. E., Gupta, S., Choi, M. et al. mobileOG-db: a manually curated database of protein families mediating the life cycle of bacterial mobile genetic elements. *Appl. Environ. Microbiol.* **88**, e00991-22 (2022).

55. Proksee. *CRISPR/Cas Finder tool*. Proksee tool version 1.1.0; underlying CRISPRCasFinder version 4.2.20. https://proksee.ca/tools/crispr_cas_finder (Accessed: 1 October 2026).

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

57. Blin, K., Shaw, S., Augustijn, H. E. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).

58. Cosentino, S., Larsen, M. V., Aarestrup, F. M. & Lund, O. PathogenFinder—distinguishing friend from foe using bacterial whole-genome sequence data. *PLoS ONE* **8**, e77302 (2013).

59. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. *JAMA* **333**, 71–74 (2024).

60. Yao, Z., Ha, U., Kim, S., Kim, J. & Kim, Y. Genetic characterization of fluoroquinolone-resistant *Escherichia coli* isolates. *J. Antimicrob. Chemother.* **59**, 196–204 (2007).

61. Nelson, M. L. & Levy, S. B. The history of the tetracyclines. *Ann. N. Y. Acad. Sci.* **879**, 19–29 (1999).

62. Fischbach, M. A., Lin, H., Zhou, L., Yu, Y., Abergel, R. J., Liu, D. R. et al. The pathogen-associated *iroA* gene cluster mediates bacterial evasion of lipocalin 2. *Nat. Chem. Biol.* **2**, 458–465 (2006).

63. Rahaman, M. M., Sarkar, M. M. H., Rahman, M. S. et al. Genomic characterization of the dominating Beta, V2 variant carrying vaccinated (Oxford–AstraZeneca) and nonvaccinated COVID-19 patient samples in Bangladesh: A metagenomics and whole-genome approach. *J. Med. Virol.* **94**, 1670–1688 (2022).

64. Russo, T. A., Carlino, U. B., Mong, A. & Jodush, S. T. Identification of genes in a novel virulence region of the CFT073 chromosome that contribute to acute cystitis. *Infect. Immun.* **76**, 250–259 (2008).

Declarations

The submitted research article “Integrated Phenotypic and Whole-Genome Characterization of a Uropathogenic Escherichia coli ST127, O6:H31 Isolate from Bangladesh” for publication in your journal of repute is a unique article, and nobody has done it earlier.

Consent to publication

Not applicable.

Data availability statement

The raw sequencing data for BCSIR SA 01 were submitted to the NCBI under BioProject accession number PRJNA1537254, BioSample accession number SAMN63953655, and Sequence Read Archive (SRA) accession numbers SRX35532930 (Experiment) and SRR41005028 (Run). (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1537254).

Conflict of interest

The authors declare that they have no conflicts of interest.

Ethical Clearance

This study was approved by the Ethical Review Committee of Jagannath University, Dhaka, Bangladesh (Approval Code: JnU/ERC/2026/64, Date: 16 September 2026).

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author's Contributions:

MAK and SPM resources, analyses, writing, and methodology; NU, TAB, MI, BG, SFC, SRN, SM, JINO: investigation, software, validation; SB guidance; SA Conceptualization, review and editing, visualization, and supervision. All authors have carefully reviewed and consented to the final version of the manuscript.