

Draft Genome Sequence Data of Multidrug-Resistant *Escherichia coli* CUK-76 Co-Harboring Class A and Class C β -Lactamases from Wastewater of India

Achhada Ujalkaur Avatsingh¹, Shilpa Sharma¹, Shilippreet Kour¹, Anvesha Bhardwaj², Prem Prashant Chaudhary^{3,*} and Nasib Singh^{1,*} 

¹ Department of Microbiology, Akal College of Basic Sciences, Eternal University, Baru Sahib, Sirmaur 173101, Himachal Pradesh, India; bs22pmb001@eternaluniversity.edu.in (A.U.A.); bs21pmb001@eternaluniversity.edu.in (S.S.); bs21pmb002@eternaluniversity.edu.in (S.K.)

² Advanced Academic Programs, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD 21218, USA; abhard10@jh.edu

³ Laboratory of Clinical Immunology and Microbiology, Epithelial Therapeutics Unit, National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, MD 20892, USA

* Correspondence: chaudharyp2@nih.gov (P.P.C.); drnasibmicro@eternaluniversity.edu.in (N.S.)

Abstract

The present study was performed to determine the antibiotic resistance genes (ARGs), virulence determinants, and mobile genetic elements in multidrug-resistant *Escherichia coli* CUK-76 isolated from wastewater in Himachal Pradesh, India. Whole genome sequencing was performed using the Illumina Miseq system, and the draft genome sequence was assembled by Unicycler v0.5.1 and annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v6.10). The bioinformatics-based prediction analysis was performed using ResFinder v4.7.2 and CARD v4.0.1 (antibiotic resistance genes), VirulenceFinder v2.0, VFDB and MGEFinder v1.0.3 (virulence determinants), PlasmidFinder v2.0.1 (plasmid sequences), MLST v2.0 (sequence type), ISFinder and TnCentral v2.0 (insertion sequences and transposons), PathogenFinder2 v0.6.0 (pathogenicity), RAST (subsystems category) and Phigaro (prophage sequences). The draft genome of *E. coli* CUK-76 strain comprised 4,607,136 bp with a GC content of 51%. Genome annotation revealed 4498 genes of which 4289 were protein-coding genes, 78 RNA genes, and 131 pseudogenes. It was related to sequence type ST949 and its predicted resistome consisted of *bla*_{CTX-M-15}, *bla*_{TEM-1B} (class A β -lactamase genes), *bla*_{EC-14} (class C β -lactamase gene), *aph*(6)-*Id*, *aph*(3'')-*Ib*, *qnrS1*, *sul2*, *tet*(A), and *dfrA14* genes. Additionally, multiple virulence genes, two plasmid sequences viz. IncFIB(K) and IncFIB(AP001918), insertion sequences, transposons and prophage sequences were detected. The genomic dataset of this strain will be a valuable resource for comparative genomic studies on *E. coli*.



Academic Editor: Pufeng Du

Received: 27 June 2026

Revised: 29 July 2026

Accepted: 4 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Dataset: The raw genome sequencing data were submitted to the NCBI Sequence Read Archive under accession number SRR37187637, BioSample accession number SAMN55285571, and BioProject accession number PRJNA1421061. The assembled genome has been deposited in NCBI under GenBank accession number JBVUNU000000000 and the genome assembly accession number GCA_055791285.1.

Dataset License: CC-BY

Keywords: whole genome sequencing; *Escherichia coli*; wastewater; antibiotic resistant genes; β -lactamases

1. Summary

Escherichia coli is an important commensal constituent of the human gut microbiota. However, some strains of this species are frequently associated with healthcare-associated infections [1]. WHO has listed *E. coli* as one of the highest-priority pathogens. *E. coli* is known to exhibit resistance to β -lactams and other classes of antibiotics and poses a severe threat to public health [2,3]. Here, we report the draft genome sequence of a multidrug-resistant *E. coli* strain isolated from municipal wastewater in Solan, Himachal Pradesh, India. Its antibiotic resistance genes (ARGs), virulence determinants, sequence type, subsystem categories, and mobile genetic elements were determined by genomic and bioinformatics analysis.

2. Data Description

E. coli CUK-76 is a multidrug-resistant strain as revealed by its phenotypic resistance to multiple antibiotic classes, viz. penicillins, cephalosporins, carbapenems, aminoglycosides, macrolides, quinolones, and sulphonamides (Supplementary Table S1). The high-quality draft genome assembly consisted of 4,607,136 bp with a GC content of 51%. Of the total 4498 predicted genes, 4420 were coding sequences, 4289 were protein-coding, 78 were RNA genes, and 131 were pseudogenes (Table 1). As indicated by BUSCO and CheckM analysis, the draft genome of *E. coli* strain CUK-76 was near-complete with 99.17% completeness and only a 0.31% contamination level (Table 1). MLST analysis confirmed the sequence type to be ST949. The subsystem annotation by RAST analysis revealed that 45 genes are associated with virulence, disease, and defense capabilities, as well as adaptive responses (Figure 1). CARD v4.0.1 and ResFinder v4.7.2 analyses revealed that the predicted resistome comprised acquired ARGs, including *bla*_{CTX-M-15}, *bla*_{TEM-1B}, *aph*(6)-*Id*, *aph*(3'')-*Ib*, *qnrS1*, *sul2*, *tet*(A), and *dfrA14* (Table 2). Additionally, the intrinsic genes and regulatory determinants coding for ATP-binding cassette (ABC) antibiotic efflux pumps, Resistance–Nodulation–Cell Division (RND) antibiotic efflux pump, and Major Facilitator Superfamily (MFS) antibiotic efflux pump, antibiotic inactivation, and decreased membrane permeability were predicted in the genome of *E. coli* strain CUK-76 (Table 2). This strain co-harbored both class A β -lactamase genes (*bla*_{CTX-M-15} and *bla*_{TEM-1B}) and a chromosomally encoded class C β -lactamase gene (*bla*_{EC-14}).

Table 1. The key genome and annotation statistics of multidrug-resistant *E. coli* strain CUK-76.

Characteristic/Annotation	Value
Draft genome size	4,607,136 bp
Number of raw reads	11,602,998
GC content (%)	51
Number of contigs	90
Contig N50/L50	141.6 kb / 13
Number of scaffolds	01
Scaffold N50/L50	4,607,136 bp / 1
Genome completeness	99.17%
Genome coverage	778.69x
Genome contamination	0.31%
Total genes	4498
Total coding sequences (CDSs)	4420

Table 1. Cont.

Characteristic/Annotation	Value
Genes (protein coding)	4289
Number of rRNA genes	1
Number of tRNA genes	69
Number of ncRNA genes	8
Number of pseudogenes	131
CRISPR arrays	2
Plasmid sequences	IncFIB(K) and IncFIB(AP001918)
Genome accession No.	JBVUNU000000000
Genome assembly accession No.	GCA_055791285.1
BioSample accession No.	SAMN55285571
BioProject accession No.	PRJNA1421061
SRA accession No.	SRR37187637

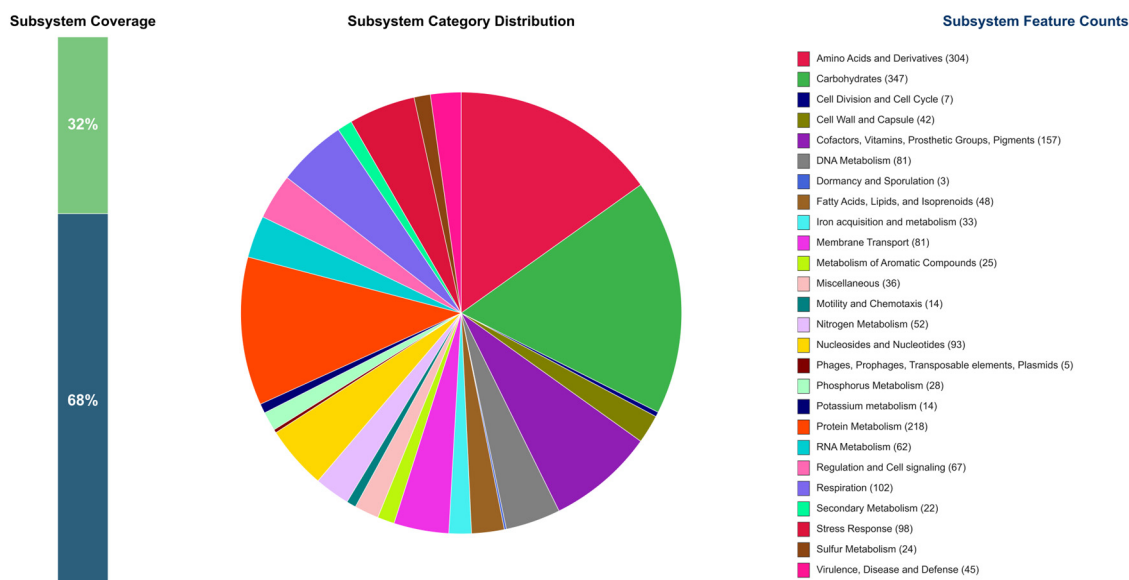


Figure 1. The subsystems category and corresponding characteristics of *E. coli* strain CUK-76 as determined by RAST annotation.

The predicted genome analysis revealed the detection of several virulence genes, of which *neuC* (capsule biosynthesis protein), *ompA* (outer membrane protein A), *espL1* (Type III secretion system (T3SS) effector), *fdeC* (adhesin), *lpfA* (long polar fimbriae), *hlyE* (hemolysin), *csgA* (major curli subunit), *entF* (enterobactin synthase), *fepA* (ferric enterobactin receptor) *terC* (tellurium ion resistance protein), and *nlpI* (lipoprotein NlpI precursor) were the prominent ones (Table 3 and Supplementary Table S3). The occurrence of virulence-associated determinants may contribute to its pathogenic potential, as is also supported by a probability score of 0.9736 predicted by PathogenFinder2. However, further experimental and clinical evidence is necessary to confirm its human pathogenicity. Genomic analysis showed the presence of two plasmid sequences, i.e., IncFIB(K) and IncFIB(AP001918). Additionally, the insertion sequences, transposable elements (Supplementary Table S2), and prophage sequences/virus orthologous group (VOGs) were also detected (Table 4).

Table 2. The antibiotic resistance genes predicted in the genome sequence of multidrug-resistant *E. coli* CUK-76.

S. No.	Gene	Gene Product/Function	Resistance Mechanism(s)	Resistance Towards	% Identity	Accession No.
Acquired antibiotic resistance genes						
1.	<i>bla</i> _{CTX-M-15}	Extended-spectrum β -lactamases (Class A)	Antibiotic inactivation	Cephalosporins and penicillins	100	AY044436
2.	<i>bla</i> _{TEM-1}	Extended-spectrum β -lactamases (Class A)	Antibiotic inactivation	Monobactams, cephalosporins and penicillins	100	AL513383
3.	<i>tet(A)</i>	MFS antibiotic efflux pump	Antibiotic efflux	Tetracyclines	100	AF534183.1
4.	<i>sul2</i>	Sulfonamide resistance	Antibiotic target replacement	Sulfonamides	100	AY055428.1
5.	<i>aph(6)-Id</i>	Aminoglycoside O-phosphotransferases (class 6)	Antibiotic inactivation	Aminoglycosides	99.64	AF024602
6.	<i>aph(3'')-Ib</i>	Aminoglycoside O-phosphotransferases (class 6)	Antibiotic inactivation	Aminoglycosides	99.64	AF313472
7.	<i>qnrS1</i>	Quinolone resistance protein (qnr)	Antibiotic target protection	Fluoroquinolones	99.64	DQ485529.1
8.	<i>dfrA14</i>	Trimethoprim-resistant dihydrofolate reductase	Reduced affinity for antibiotics	Trimethoprim	100	KF921535
Intrinsic antibiotic resistance genes/regulatory determinants						
1.	<i>bla</i> _{EC-14}	AmpC β -lactamases (Class C)	Antibiotic inactivation	Cephalosporins	98.67	WP_063860233.1
2.	<i>mdtE</i>	Resistance–Nodulation–Cell Division (RND) antibiotic efflux pump	Antibiotic efflux	Macrolides, fluoroquinolones and penicillins	100	AP009048.1
3.	<i>acrB</i>	RND antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones, cephalosporins, glycylicyclines, penicillins, tetracyclines, rifamycin, and phenicols	100	U00096.3
4.	<i>acrA</i>	RND antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones, cephalosporins, glycylicyclines, penicillins, tetracyclines, rifamycin, and phenicols	100	GCF_000005845.2
5.	<i>acrE</i>	RND antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones, cephalosporins, penicillins	100	U00096

Table 2. Cont.

S. No.	Gene	Gene Product/Function	Resistance Mechanism(s)	Resistance Towards	% Identity	Accession No.
6.	<i>cpxA</i>	RND antibiotic efflux pump	Antibiotic efflux	Aminoglycosides and aminocoumarins	100	BA000007.3
7.	<i>mdtA</i> , <i>mdtB</i> , <i>mdtC</i>	RND antibiotic efflux pump	Antibiotic efflux	Aminocoumarins	99.52	U00096
8.	<i>gadX</i>	RND antibiotic efflux pump	Antibiotic efflux	Macrolides, fluoroquinolones, and penicillins	98.18	AP009048.1
9.	<i>mdtF</i>	RND antibiotic efflux pump	Antibiotic efflux	Macrolides, fluoroquinolones, and penicillins	98.18	U00096
10.	<i>marA</i>	RND antibiotic efflux pump, general bacterial porin with reduced permeability to beta-lactams	Antibiotic efflux, reduced permeability to antibiotics	Fluoroquinolones, monobactams, carbapenems, cephalosporins, glycylicyclines, penicillins, tetracyclines, rifamycin and phenicols	100	AP009048.1
11.	<i>acrS</i>	RND antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones, cephalosporins, glycylicyclines, penicillins, tetracyclines, rifamycin, and phenicols	99.55	U00096
12.	<i>acrF</i>	RND antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones, cephalosporins, glycylicyclines, penicillins, tetracyclines, rifamycin, and phenicols	99.55	U00096
13.	<i>baeR</i>	RND antibiotic efflux pump	Antibiotic efflux	Aminoglycosides and aminocoumarins	95.82	AP009048.1
14.	<i>msbA</i>	ATP-binding Cassette (ABC) antibiotic efflux pump	Antibiotic efflux	Nitroimidazoles	100	U00096.3

Table 2. Cont.

S. No.	Gene	Gene Product/Function	Resistance Mechanism(s)	Resistance Towards	% Identity	Accession No.
15.	<i>tolC</i>	ABC antibiotic efflux pump, MFS antibiotic efflux pump, and RND antibiotic efflux pump	Antibiotic efflux	Macrolides, fluoroquinolones, aminoglycosides, carbapenems, cephalosporins, glycyclcyclines, penicillins, tetracyclines, peptide antibiotics, aminocoumarins, rifamycin, and phenicols	100	FJ768952
16.	<i>mdtG</i>	Major Facilitator Superfamily (MFS) antibiotic efflux pump	Antibiotic efflux	Phosphonic acid antibiotics	100	CP000800.1
17.	<i>emrR</i>	MFS antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones	100	U00096.3
18.	<i>emrB</i>	MFS antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones	100	U00096
19.	<i>emrY</i>	MFS antibiotic efflux pump	Antibiotic efflux	Tetracyclines	100	D78168
20.	<i>emrK</i>	MFS antibiotic efflux pump	Antibiotic efflux	Tetracyclines	100	D78168
21.	<i>mdtM</i>	MFS antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones, lincosamides, nucleoside antibiotics, and phenicols	98.04	U00096.3
22.	<i>emrA</i>	MFS antibiotic efflux pump	Antibiotic efflux	Fluoroquinolones	99.74	AP009048
23.	<i>mdfA</i>	MFS antibiotic efflux pump	Antibiotic efflux	Tetracyclines and phenicols	96.34	JQ394987

Table 3. The virulence genes predicted in the genome sequence of multidrug-resistant *E. coli* CUK-76.

S. No.	Gene	Gene Product/Function	Accession No.	Position in Scaffold	Coverage (%)	Identity (%)	Predicted by
1.	<i>capU</i>	Hexosyltransferase homolog	CU928145	12,132–13,220	100	99.63	VirulenceFinder v2.0
2.	<i>csgA</i>	Curlin major subunit CsgA	CP069646	149,418–149,873	100	98.9	VirulenceFinder v2.0
3.	<i>csgB</i>	Minor curlin subunit precursor curli nucleator protein CsgB	NP_460114	1,179,022–1,179,477	99.78	83.81	VFDB
4.	<i>csgD</i>	DNA-binding transcriptional regulator CsgD [curli fibers/thin aggregative fimbriae	NP_460113	1,180,231–1,180,881	100	81.11	VFDB

Table 3. Cont.

S. No.	Gene	Gene Product/Function	Accession No.	Position in Scaffold	Coverage (%)	Identity (%)	Predicted by
5.	<i>csgF</i>	Curli production assembly/transport protein	NP_460111	1,181,300–1,181,716	99.04	81.23	VFDB
6.	<i>csgG</i>	Curli production assembly/transport protein CsgG	NP_460110	1,181,743–1,182,576	100	82.97	VFDB
7.	<i>entA</i>	23-dihydro-23-dihydroxybenzoate dehydrogenase [Enterobactin (VF0228)]	NP_752614	371,791–372,537	100	95.98	VFDB
8.	<i>entB</i>	Isochorismatase [Enterobactin (VF0228)]	NP_752613	372,537–373,394	100	97.44	VFDB
9.	<i>entC</i>	Isochorismate synthase 1 [Enterobactin (VF0228)]	NP_752611	375,028–376,215	100	97.56	VFDB
10.	<i>entD</i>	Phosphopantetheinyl transferase component of enterobactin synthase multienzyme complex	NP_752599	390,859–391,629	100	94.03	VFDB
11.	<i>entE</i>	23-dihydroxybenzoate-AMP ligase component of enterobactin synthase multienzyme complex	NP_752612	373,408–375,018	100	95.9	VFDB
12.	<i>entF</i>	Enterobactin synthase multienzyme complex component ATP-dependent	NP_752604	383,078–386,931	99.28	95.61	VFDB
13.	<i>entS</i>	Enterobactin exporter iron-	NP_752609	377,538–378,788	100	95.52	VFDB
14.	<i>espL1</i>	Type III secretion system effector espL1	NP_288154	2,216,858–2,218,756	100	94.79	VFDB
15.	<i>espR1</i>	Type III secretion system effector espR1	NP_287686	781,474–782,529	83.81	94.13	VFDB
16.	<i>espX1</i>	Type III secretion system effector EspX1	NP_285716	1,812,054–1,813,475	100	94.44	VFDB
17.	<i>espX5</i>	Type III secretion system effector EspX5	NP_290699	120,043–121,335	100	95.59	VFDB
18.	<i>fdeC</i>	Adhesin FdeC	YP_002390132	2,543,928–2,548,169	99.76	92.35	VFDB
19.	<i>fepA</i>	Ferrienterobactin outer membrane transporter	NP_752600	388,594–390,834	100	96.79	VFDB
20.	<i>fepB</i>	Ferrienterobactin ABC transporter periplasmic binding protein	NP_752610	376,578–377,534	100	97.28	VFDB
21.	<i>fepC</i>	Ferrienterobactin ABC transporter ATPase	NP_752606	380,889–381,704	100	96.94	VFDB
22.	<i>fepD</i>	Ferrienterobactin ABC transporter permease	NP_752608	378,887–379,903	100	95.67	VFDB
23.	<i>fepG</i>	Iron-enterobactin ABC transporter permease	NP_752607	379,900–380,892	100	94.56	VFDB

Table 3. Cont.

S. No.	Gene	Gene Product/Function	Accession No.	Position in Scaffold	Coverage (%)	Identity (%)	Predicted by
24.	<i>fes</i>	Enterobactin/ferric enterobactin esterase	NP_752602	387,149–388,351	100	96.76	VFDB
25.	<i>gspC</i>	General secretion pathway protein C	YP_404599	3,443,383–3,444,213	100	95.07	VFDB
26.	<i>gspD</i>	General secretion pathway protein D	YP_404600	3,444,453–3,446,303	100	97.3	VFDB
27.	<i>gspE</i>	General secretion pathway protein E	YP_404601	3,446,303–3,447,796	100	96.25	VFDB
28.	<i>gspF</i>	General secretion pathway protein F	YP_404602	3,447,796–3,448,995	100	96.33	VFDB
29.	<i>gspG</i>	General secretion pathway protein G	YP_404603	3,449,036–3,449,491	100	95.61	VFDB
30.	<i>gspH</i>	General secretion pathway protein H	YP_404604	3,449,528–3,450,058	100	95.86	VFDB
31.	<i>gspI</i>	General secretion pathway protein I	YP_404605	3,450,055–3,450,426	100	95.16	VFDB
32.	<i>gspJ</i>	General secretion pathway protein J	YP_404606	3,450,459–3,451,028	100	96.49	VFDB
33.	<i>gspK</i>	General secretion pathway protein K	YP_404607	3,451,025–3,452,002	100	96.11	VFDB
34.	<i>gspL</i>	General secretion pathway protein L	YP_404608	3,452,317–3,453,177	100	97.56	VFDB
35.	<i>gspM</i>	General secretion pathway protein M	YP_404609	3,453,293–3,453,715	100	94.09	VFDB
36.	<i>hlyE</i>	Avian <i>E. coli</i> haemolysin	ECU57430	39,018–39,935	100	99.13	VirulenceFinder v2.0
37.	<i>lpfA</i>	Long polar fimbriae	KC207123	29,904–30,476	100	100	VirulenceFinder v2.0
38.	<i>neuC</i>	Polysialic acid capsule biosynthesis protein	NZ_NJSA01000011	38,723–39,886	100	100	VirulenceFinder v2.0
39.	<i>nlpI</i>	Lipoprotein NlpI precursor	CP000243	40,998–41,882	100	99.77	VirulenceFinder v2.0
40.	<i>ompA</i>	Outer membrane protein A	AAF37887	1,252,095–1,253,147	100	94.3	VFDB
41.	<i>terC</i>	Tellurium ion resistance protein	CP007491	94,086–94,799	100	100	VirulenceFinder v2.0
42.	<i>yagV/ecpE</i>	<i>E. coli</i> common pilus chaperone EcpE	NP_286006	2,534,645–2,535,398	99.74	96.68	VFDB
43.	<i>yagW/ecpD</i>	Polymerized tip adhesin of ECP fibers	NP_286007	2,535,322–2,536,965	100	98.6	VFDB
44.	<i>yagX/ecpC</i>	<i>E. coli</i> common pilus usher EcpC	NP_286008	2,536,955–2,539,480	100	98.65	VFDB
45.	<i>yagY/ecpB</i>	<i>E. coli</i> common pilus chaperone EcpB	NP_286009	2,539,506–2,540,174	100	98.66	VFDB

Table 3. Cont.

S. No.	Gene	Gene Product/Function	Accession No.	Position in Scaffold	Coverage (%)	Identity (%)	Predicted by
46.	<i>yagZ/ecpA</i>	<i>E. coli</i> common pilus structural subunit EcpA	NP_286010	2,540,232–2,540,819	100	99.32	VFDB
47.	<i>yehA</i>	Outer membrane lipoprotein, YHD fimbriael cluster	CP042934	151,133–152,165	99.71	98.06	VirulenceFinder v2.0
48.	<i>yehB</i>	Usher, YHD fimbriael cluster	CP042934	152,181–154,661	100	97.86	VirulenceFinder v2.0
49.	<i>yehC</i>	Chaperone, YHD fimbriael cluster	CP042934	154,677–155,351	100	98.52	VirulenceFinder v2.0
50.	<i>yehD</i>	Major pilin subunit, YHD fimbriael cluster	CP042934	155,432–155,974	100	98.71	VirulenceFinder v2.0

Table 4. Supportive characteristics of draft genome of *E. coli* strain CUK-76 in relation to transposable elements, insertion sequences, and prophage sequences (Virus orthologous group, VOGs).

Category	Genetic Elements	Accession No./Description
Insertion sequences (IS)	IS100, IS100kyp, IS100L, IS100X, IS1203, IS1203E, IS1230B, IS1A, IS1D, IS1F, IS1G, IS1H, IS1R, IS1S, IS1SD, IS1 × 1, IS1 × 2, IS1 × 3, IS1 × 4, IS21, IS3, IS30, IS30D, IS30H, IS3411, IS3F, IS3H, IS5, IS5B, IS5D, IS609, IS629, IS630, IS640, IS682, IS911, ISAcsp1, ISArsp6, ISAs26, ISBrsa1, ISEal1, ISEc1, ISEc16, ISEc17, ISEc23, ISEc25, ISEc26, ISEc33, ISEc38, ISEc41, ISEc43, ISEc44, ISEc47, ISEc5, ISEc52, ISEc62, ISEc97, ISEhe1, ISEic2, ISEsp1, ISKox1, ISKpn14, ISKpn26, ISKpn33, ISKpn41, ISKpn47, ISKpn69, ISLad1, ISPa26, ISPa38, ISPa40, ISPa41, ISPa52, ISPeat2, ISPsy2, ISRor6, ISSba14, ISSen14, ISSen3, ISSen6, ISSen8, ISSen9, ISShdy4, ISShes11, ISSlo2, ISSma18, ISSma5, ISSma7, ISSma9, ISSpu2, ISStes1, ISSty2, ISYru1, MITEEc1, MITESen1, and MITEYpe1	The detailed characteristics of insertion sequences are provided in Supplementary Table S2
Transposable elements (transposons)	Tn4430, Tn5393, TnAs1, TnAs2, TnAs3, TnEc1, TnEc3, TnShfr1	Transposons were predicted by ISFinder and TnCentral. Their details are provided in Supplementary Table S2.

Table 4. Cont.

Category	Genetic Elements	Accession No./Description
Prophage sequences/Virus orthologous group (VOGs)	VOG0583, VOG0582, VOG4736, VOG5357, VOG4630, VOG0577, VOG0356, VOG0221, VOG10343, VOG0574, VOG4691, VOG0573, VOG4645, VOG4565, VOG0765, VOG0569, VOG0568, VOG0567, VOG0764, VOG0566, VOG3897, VOG4556, VOG8502, VOG4814, VOG2188, VOG0588, VOG4694, VOG0584, VOG2187, VOG2186, VOG3379, VOG2215, VOG5628, VOG0275, VOG6627, VOG6451, VOG6452, VOG6453, VOG0585, VOG0586, VOG0979, VOG0764, VOG2217, VOG0614, VOG6912, VOG0237, VOG0571, VOG0572, VOG4550, VOG9196, VOG0227, VOG0356, VOG0577, VOG4630, VOG4545, VOG4736, VOG0582, VOG0583, VOG6520, VOG7686, VOG7437, VOG9667, VOG4573, VOG6001, VOG3434, VOG1887, VOG0838, VOG3327, VOG2791, VOG3328, VOG6672, VOG2792	Belongs to family <i>Myoviridae</i>
	VOG5994, VOG8647, VOG2124, VOG2341, VOG0251, VOG4693, VOG0298, VOG5092, VOG4565, VOG3405, VOG8908, VOG7882, VOG4771, VOG0051, VOG4361, VOG0833, VOG4672, VOG4364, VOG4438, VOG0839, VOG0840, VOG4563, VOG0842, VOG0843, VOG0847, VOG0848, VOG0765, VOG0299, VOG4618, VOG0590, VOG0796, VOG6488, VOG1518, VOG6152, VOG4598, VOG0855, VOG10343	Belongs to family <i>Siphoviridae</i>

Multidrug efflux pumps and other intrinsic resistance mechanisms are potentially resistant to tetracycline, azithromycin, and imipenem. It is also possible that *E. coli* CUK-76 may harbor previously uncharacterized resistance genes or novel gene variants that are not listed in the current CARD and ResFinder database. Furthermore, the stringent detection thresholds applied in the CARD and ResFinder databases for gene identification may also lead to non-detection of those ARGs that are either highly divergent or have truncated sequences. The dataset presented in this study provides a detailed genomic elucidation of *E. coli* CUK-76 belonging to sequence type ST949 and harboring β -lactamase-encoding genes and other resistance determinants. Furthermore, the detection of virulence determinants and a high pathogenicity score indicates its potential pathogenicity and the possible prevalence of other multidrug-resistant bacteria in the wastewater of the investigated geographical site. Our study has limitations as it is confined to a single geographical location and lacks temporal and spatial replication. However, the findings of this report warrant further studies on a larger set of bacterial isolates from a broader geographical area to unravel the potential role of wastewater-inhabiting bacteria in the dissemination of antibiotic resistance.

3. Methods

3.1. Sample Collection and Isolation of *E. coli* Strain CUK-76

E. coli CUK-76 was isolated from municipal wastewater in Solan (coordinates 30.925161° N, 76.795547° E), Himachal Pradesh, India. About 1 L of wastewater sample was collected in sterile screw-capped polystyrene bottles and transported to the laboratory in an insulated box for microbiological analysis [4]. About 1.0 mL of the sample was plated on MacConkey agar (HiMedia Laboratories, Mumbai, India), incubated at 35–37 °C for 24 h, and pure cultures were obtained as described earlier [5,6]. The strain was characterized by the HiIMViC biochemical test kit (HiMedia Laboratories, Mumbai, India) and 16S rRNA sequencing (GenBank accession PZ134701). The antibiotic susceptibility of *E. coli* CUK-76 was determined against different classes of antibiotics by the Kirby–Bauer disk diffusion method in accordance with the standard guidelines of the Clinical Laboratory Standards Institute [7] and Drieux et al. [8]. The results were interpreted as resistant, intermediate, and susceptible according to CLSI breakpoints.

3.2. DNA Isolation, Genome Sequencing, Assembly, and Annotation

The bacterial isolate was grown in nutrient broth at 37 °C for 24 h, and its genomic DNA was extracted using an Xploreagen DNA extraction kit (Xploreagen, Bengaluru, India). The quantity of the extracted DNA was measured using a Qubit 3 fluorometer (Thermo Fisher Scientific, Newark, DE, USA) to ensure sufficient concentration for sequencing [9]. For library preparation, 100 ng of genomic DNA was enzymatically fragmented to obtain DNA fragments of approximately 200–300 bp in length. These fragments were then subjected to end repair and 3' adenylation. Loop adapters were ligated to the prepared fragments, followed by cleavage using an uracil-specific excision reagent enzyme. The adapter-ligated DNA was purified using AMPure beads and enriched through six cycles of PCR amplification using NEBNext Ultra II Q5 master mix and Illumina-specific primers. The final sequencing library was again quantified using the Qubit 3 fluorometer (Thermo Fisher Scientific, DE, USA), and fragment size distribution was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The prepared libraries were then sequenced on the Illumina Miseq system using 150 bp paired-end reads [6,9–11].

The quality of the raw sequencing reads was first evaluated using FastQC v0.12.1 [12]. High-quality reads were then assembled into a draft genome using Unicycler v0.5.1 [13]. Genome completeness was assessed using CheckM analysis v1.2.5 [14] and BUSCO

v6.0.0 [15]. Genome coverage was estimated based on the sequencing reads relative to the assembled genome using BBTools: BBNorm [16]. The generated contigs were further arranged and oriented using the scaffolding tool Multi-CSAR v1.0 [17]. Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v6.10) to identify coding sequences and other genomic features [18]. Antibiotic resistance genes were predicted in the genome sequence using the Comprehensive Antibiotic Resistance Database (CARD v4.0.1) [19] and ResFinder v4.7.2 [20]. Only “perfect hits” (100% identity and coverage) and “strict hits” ($\geq 95\%$ identity and $\geq 80\text{--}90\%$ coverage) were considered for analysis. The plasmid sequences and virulence determinants were predicted by PlasmidFinder v2.0.1 [21] and VirulenceFinder v2.0 [22] and the Virulence Factor Database (VFDB) [23], respectively. Mobile genetic elements (insertion sequences and transposable elements) were detected using MGEfinder v1.0.3 [24], ISFinder [25], and TnCentral v2.0 [26] databases. The sequence type of the strain was determined by MLST v2.0 [27], whereas the pathogenic potential was determined by PathogenFinder v2.0.6 [28]. Subsystem categories, gene functions, and metabolic networks were annotated by Rapid Annotation using Subsystems Technology (RAST) annotation [29]. Prophage sequences or Viral Orthologous Groups (VOGs) in the genome sequence were determined by the Phigaro database [30]. All analyses with various databases were performed with default settings unless stated otherwise.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/data11080197/s1>, Table S1: Antibiotic susceptibility profile of *E. coli* strain CUK-76; Table S2: List of mobile genetic elements (insertion sequences and transposons) predicted in the draft genome of *E. coli* strain CUK-76; Table S3: List of virulence genes and their annotation characteristics predicted from the genome sequence of *E. coli* strain CUK-76.

Author Contributions: Conceptualization, N.S. and P.P.C.; methodology, A.U.A., S.K. and S.S.; formal analysis, N.S. and S.K.; investigation, N.S., S.K., S.S. and A.U.A.; resources, N.S.; writing—original draft preparation, A.U.A., S.K., S.S., A.B., P.P.C. and N.S.; writing—review and editing, N.S., A.U.A., S.K., S.S., A.B. and P.P.C.; supervision, N.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome sequence data of *E. coli* strain CUK-76 have been submitted to the NCBI Sequence Read Archive (SRA) under accession number SRR37187637, BioSample accession number SAMN55285571, and BioProject accession number PRJNA1421061. The assembled genome has been deposited in NCBI under Genome accession number JBVUNU000000000 and genome assembly accession number GCA_055791285.1.

Acknowledgments: The authors are thankful to Eternal University for providing the necessary facilities and support for the research study. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. The authors would like to thank the team of Biokart India Pvt. Ltd., Bengaluru (India) for providing the WGS services.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AMR	Antimicrobial resistance
MDR	Multidrug resistant
ARGs	Antibiotic resistance genes
WGS	Whole genome sequencing
CDS	Coding DNA Sequences
SRA	Sequence Read Archive
CLSI	Clinical Laboratory Standards Institute
CARD	Comprehensive Antibiotic Resistance Database
RAST	Rapid Annotation using Subsystems Technology
RND	Resistance–Nodulation–Cell Division
ABC	ATP-binding Cassette
MFS	Major Facilitator Superfamily

References

1. Kerek, A.; Roman, I.; Szabo, A.; Kovacs, D.; Kardos, G.; Kovacs, L.; Jerzsele, Á. Antibiotic resistance genes in *Escherichia coli*-literature review. *Crit. Rev. Microbiol.* **2026**, *52*, 1–35. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Sati, H.; Carrara, E.; Savoldi, A.; Hansen, P.; Garlasco, J.; Campagnaro, E.; Boccia, S.; Castillo-Polo, J.A.; Magrini, E.; Garcia-Vello, P.; et al. WHO Bacterial Priority Pathogens List Advisory Group. The WHO Bacterial Priority Pathogens List 2024: A prioritisation study to guide research, development, and public health strategies against antimicrobial resistance. *Lancet Infect. Dis.* **2025**, *25*, 1033–1043. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Zhong, Y.; Teo, J.Q.M.; Guo, S.; Schlundt, J.; Kwa, A.L.H.; Ong, R.T.H. Characterization of mobile resistance elements in extended-spectrum β -lactamase producing gram-negative bacteria from aquatic environment. *Sci. Total Environ.* **2025**, *978*, 179353. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Avatsingh, A.U.; Sharma, S.; Kour, S.; Chaudhary, P.P.; Singh, N. Draft genome sequence of multidrug-resistant *Enterobacter bugandensis* CUK-31 from wastewater of Himachal Pradesh, India. *Microbiol. Resour. Announc.* **2026**, *15*, e00175-26. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Kour, S.; Sharma, S.; Avatsingh, A.U.; Chaudhary, P.P.; Singh, N. Draft genome sequence of multidrug-resistant *Escherichia coli* strain ASK-30 from wastewater in northern India. *Microbiol. Resour. Announc.* **2026**, *15*, e00009-26. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Sharma, S.; Avatsingh, A.U.; Kour, S.; Chaudhary, P.P.; Singh, N. Whole genome sequencing of multidrug-resistant *Escherichia coli* strain BSS-62 from wastewater. *Microbiol. Resour. Announc.* **2026**, *15*, e00378-26. [\[CrossRef\]](#) [\[PubMed\]](#)
7. CLSI M100-S30; Performance Standards for Antimicrobial Susceptibility Testing. 31st ed. Clinical and Laboratory Standards Institute: Wayne, PA, USA, 2021.
8. Drieux, L.; Brossier, F.; Sougakoff, W.; Jarlier, V. Phenotypic detection of extended-spectrum beta-lactamase production in Enterobacteriaceae: Review and bench guide. *Clin. Microbiol. Infect.* **2008**, *14*, 90–103. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Kour, S.; Sharma, S.; Avatsingh, A.U.; Chaudhary, P.P.; Singh, N. Draft genome sequence data of multidrug-resistant *Pseudomonas aeruginosa*, strain ASK-80. *Data* **2026**, *11*, 96. [\[CrossRef\]](#)
10. Barbian, K.; Bruno, D.; Sykora, L.; Ricklefs, S.; Chaudhary, P.P.; Beare, P.A.; Myles, I.A.; Martens, C.M. *De novo* assembly of *Roseomonas mucosa* isolates from healthy human volunteers used to treat atopic dermatitis. *Microbiol. Resour. Announc.* **2023**, *12*, e00520-23. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Barbian, K.; Bruno, D.; Sykora, L.; Ricklefs, S.; Chaudhary, P.P.; Beare, P.A.; Myles, I.A.; Martens, C.M. *De novo* assembly of *Roseomonas mucosa* isolated from patients with atopic dermatitis. *Microbiol. Resour. Announc.* **2023**, *12*, e00521-23. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 1 June 2025).
13. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome. Res.* **2015**, *25*, 1043–1055. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Tegenfeldt, F.; Kuznetsov, D.; Manni, M.; Berkeley, M.; Zdobnov, E.M.; Kriventseva, E.V. OrthoDB and BUSCO update: Annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.* **2025**, *53*, D516–D522. [\[CrossRef\]](#)
16. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. BBTools Software Package. 2014. Available online: <https://bbmap.org> (accessed on 8 June 2025).

17. Liu, S.C.; Ju, Y.R.; Lu, C.L. Multi-CSAR: A web server for scaffolding contigs using multiple reference genomes. *Nucleic Acids Res.* **2022**, *50*, W500–W509. [[CrossRef](#)] [[PubMed](#)]
18. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [[CrossRef](#)] [[PubMed](#)]
19. Alcock, B.P.; Huynh, W.; Chalil, R.; Smith, K.W.; Raphenya, A.R.; Wlodarski, M.A.; Edalatmand, A.; Petkau, A.; Syed, S.A.; Tsang, K.K.; et al. CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **2023**, *51*, D690–D699. [[CrossRef](#)] [[PubMed](#)]
20. Florensa, A.F.; Kaas, R.S.; Clausen, P.T.L.C.; Aytan-Aktug, D.; Aarestrup, F.M. ResFinder—An open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb. Genom.* **2022**, *8*, e000748. [[CrossRef](#)] [[PubMed](#)]
21. Carattoli, A.; Zankari, E.; García-Fernández, A.; Voldby Larsen, M.; Lund, O.; Villa, L.; Møller Aarestrup, F.; Hasman, H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **2014**, *58*, 3895–3903. [[CrossRef](#)] [[PubMed](#)]
22. Joensen, K.G.; Scheut, F.; Lund, O.; Hasman, H.; Kaas, R.S.; Nielsen, E.M.; Aarestrup, F.M. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **2014**, *52*, 1501–1510. [[CrossRef](#)] [[PubMed](#)]
23. Chen, L.; Yang, J.; Yu, J.; Yao, Z.; Sun, L.; Shen, Y.; Jin, Q. VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* **2005**, *33*, D325–D328. [[CrossRef](#)] [[PubMed](#)]
24. Johansson, M.H.K.; Bortolaia, V.; Tansirichaiya, S.; Aarestrup, F.M.; Roberts, A.P.; Petersen, T.N. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: Mobile Element Finder. *J. Antimicrob. Chemother.* **2021**, *76*, 101–109. [[CrossRef](#)] [[PubMed](#)]
25. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M. ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **2006**, *34*, D32–D36. [[CrossRef](#)] [[PubMed](#)]
26. Ross, K.; Varani, A.M.; Snesrud, E.; Huang, H.; Alvarenga, D.O.; Zhang, J.; Wu, C.; McGann, P.; Chandler, M. TnCentral: A prokaryotic transposable element database and web portal for transposon analysis. *MBio* **2021**, *12*, e0206021. [[CrossRef](#)] [[PubMed](#)]
27. Ibarz Pavón, A.B.; Maiden, M.C. Multilocus sequence typing. *Methods Mol. Biol.* **2009**, *551*, 129–140. [[CrossRef](#)] [[PubMed](#)]
28. Ferrer Florensa, A.; Almagro Armenteros, J.J.; Kaas, R.S.; Clausen, P.T.L.C.; Nielsen, H.; Rost, B.; Aarestrup, F.M. Whole-genome prediction of bacterial pathogenic capacity on novel bacteria using protein language models with PathogenFinder2. *Bioinformatics* **2026**, *42*, btag129. [[CrossRef](#)] [[PubMed](#)]
29. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [[CrossRef](#)] [[PubMed](#)]
30. Starikova, E.V.; Tikhonova, P.O.; Prianichnikov, N.A.; Rands, C.M.; Zdobnov, E.M.; Ilina, E.N.; Govorun, V.M. Phigaro: High-throughput prophage sequence annotation. *Bioinformatics* **2020**, *36*, 3882–3884. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.