

UC Davis

UC Davis Previously Published Works

Title

Draft genome sequences of five multidrug-resistant *Escherichia coli* strains isolated from vegetable samples in Bangladesh.

Permalink

<https://escholarship.org/uc/item/8qs2x47d>

Journal

Microbiology Resource Announcements, 13(1)

Authors

Pramanik, Pritom

Rana, Md

Ramasamy, Srinivasan

et al.

Publication Date

2024-01-17

DOI

10.1128/mra.00982-23

Peer reviewed

# Draft genome sequences of five multidrug-resistant *Escherichia coli* strains isolated from vegetable samples in Bangladesh

Md. Saiful Islam,<sup>1,2</sup> Pritom Kumar Pramanik,<sup>1</sup> Md. Liton Rana,<sup>1</sup> Srinivasan Ramasamy,<sup>3</sup> Pepijn Schreinemachers,<sup>4</sup> Ricardo Oliva,<sup>3</sup> Md. Tanvir Rahman<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 3.

**ABSTRACT** Reports indicate that vegetables are becoming a source of multidrug-resistant (MDR) bacteria, including *Escherichia coli*. Here, we present genome sequences of five MDR *E. coli* strains to assist future genomic analysis of this bacterium. These *E. coli* strains were isolated from vegetable samples of different gardening systems in Dhaka, Bangladesh.

**KEYWORDS** vegetables, gardening systems, *Escherichia coli*, whole-genome sequencing, MDR, antibiotic resistance genes, virulence factor genes, public health, Bangladesh

The overuse of antibiotics has led to the development of antibiotic resistance, giving rise to a wide range of multidrug-resistant (MDR) strains that present a substantial global health threat (1, 2). The versatile and adaptable capabilities of *Escherichia coli* make it very commonly found in food, water, and soil (3). The presence of MDR *Escherichia coli* in vegetables poses a public health issue by transferring to the human population via the food supply chain.

Between September 2022 and March 2023, fresh vegetable samples (Table 1) were collected from different gardening systems from Dhaka (23.8105°N, 90.3372°E) district of Bangladesh and transported to the laboratory (24.7245°N, 90.4372°E). The samples were processed following the procedures of the previous study (4). Briefly, vegetable samples were aseptically chopped, weighed (50 g), and placed into a sterile polyethylene stomacher bag containing 200 mL of buffered peptone water and macerated for 5 min at 230 rpm in a Stomacher 400 circulator (Seward Ltd., London, UK). The processed samples were then incubated at 37°C for 24 h, spread on eosin methylene blue agar plates, and incubated at 37°C overnight. The resulting colonies were subjected to Gram staining and biochemical tests (indole, methyl red, Voges-Proskauer, citrate utilization, and sugar fermentation tests) to isolate *E. coli* (5). *E. coli* identification was then performed using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (6). For this study, five MDR *E. coli* strains (MTR\_DSS\_V097, MTR\_DSS\_V098, MTR\_DNS\_V20, MTR\_DSR\_V504, and MTR\_DNS\_V06) were chosen and incubated in nutrient broth (HiMedia, India) at 37°C overnight. DNA was subsequently extracted from the cultured broth using a DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany). The DNA library was prepared using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). Genome sequencing was carried out on the Illumina NextSeq2000 platform, which generated paired-end reads with a length of 2 × 150 bp. The genome assembly was conducted using Unicycler v.0.4.9 (7), following a preliminary step of trimming the raw paired-end reads (Table 1) with Trimmomatic v.0.39 (8) (leading: 20, sliding window: 4:20:20, trailing: 20, and minlen = 36) to remove Illumina adapters, known Illumina artifacts, and phiX reads from the data set. Quality assessment was performed using FastQC v.0.11.7 (9). The annotation of the genome was done using PGAP v.6.6 (10). In our assembled genome, sequence types were predicted by MLST v.2.0 (11), pathogenicity

**Editor** David Rasko, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Md. Tanvir Rahman, tanvirahman@bau.edu.bd.

The authors declare no conflict of interest.

See the funding table on p. 3.

**Received** 15 October 2023

**Accepted** 9 November 2023

**Published** 14 December 2023

Copyright © 2023 Islam et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

**TABLE 1** Genomic features of the sequenced five multidrug-resistant *Escherichia coli* strains isolated from different vegetable samples from urban and peri-urban gardening systems in Dhaka, Bangladesh

	MTR_DSS_V097	MTR_DSS_V098	MTR_DNS_V20	MTR_DSR_V504	MTR_DNS_V06
<b>Samples (scientific name)</b>	Red amaranth ( <i>Amaranthus cruentus</i> )	Radish leaves ( <i>Raphanus sativus</i> )	Green chillies ( <i>Capsicum frutescens</i> )	Tomato ( <i>Solanum lycopersicum</i> )	Malabar spinach ( <i>Basella alba</i> )
BioProject accession no.	PRJNA1019895				
BioSample accession no.	SAMN37503130	SAMN37503131	SAMN37503129	SAMN37503132	SAMN37503128
SRA accession no.	SRR26156588	SRR26156587	SRR26156586	SRR26156585	SRR26156584
GenBank accession no.	JAVTVL0000000000	JAVTVM0000000000	JAVTVN0000000000	JAVTVO0000000000	JAVTVP0000000000
Total length (bp)	4,745,385	4,773,716	4,331,091	4,778,128	4,896,184
Coverage (x)	55.0404	73.5677	80.0333	60.0396	69.8073
No. of contigs	105	142	125	134	128
No. of reads	1,043,863	1,454,216	1,424,402	1,146,569	1,461,125
Contig L50	7	14	19	11	11
Contig N50 (bp)	205,460	147,301	77,403	152,699	137,565
GC content (%)	50.67	50.69	50.93	50.69	50.80
Genes (total)	4,630	4,710	4,225	4,718	4,862
Coding Sequences (CDSs) (total)	4,548	4,627	4,145	4,627	4,773
Genes (coding)	4,447	4,476	4,030	4,474	4,589
CDSs (with protein)	4,447	4,476	4,030	4,474	4,589
No. of RNA genes (tRNAs, rRNAs, and ncRNAs)	82 (71, 2, and 9)	83 (72, 2, and 9)	80 (69, 2, and 9)	91 (78, 2, and 11)	89 (77, 3, and 9)
Pseudo genes (total)	101	151	115	153	184
Sequence types (STs)	441	2	1162	2	398
Pathogenicity index	0.944	0.932	0.945	0.933	0.937
No. of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays (no. of genes) (list of genes)	2 (10) ( <i>csa3</i> , <i>cas2</i> , <i>cas1</i> , <i>cas6e</i> , <i>cas5</i> , <i>cas7</i> , <i>cas8e</i> , <i>cse2gr11</i> , <i>cas7</i> , 3 (8) ( <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , <i>csa3</i> , <i>cas5</i> , <i>cas6e</i> , <i>cas3</i> , and <i>cas1</i> ))	2 (10) ( <i>WYL</i> , <i>cas3</i> , <i>cas8e</i> , <i>cse2gr11</i> , <i>cas7</i> , 3 (8) ( <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , <i>csa3</i> , <i>cas8e</i> , <i>cas3</i> , and <i>WYL</i> ))	2 (10) ( <i>WYL</i> , <i>cas3</i> , <i>cas8e</i> , <i>cse2gr11</i> , <i>cas7</i> , 3 (12) ( <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , <i>csa3</i> , <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , and <i>csa3</i> ))	2 (10) ( <i>WYL</i> , <i>cas3</i> , <i>cas8e</i> , <i>cse2gr11</i> , <i>cas7</i> , 3 (12) ( <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , <i>csa3</i> , <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , and <i>csa3</i> ))	2 (10) ( <i>WYL</i> , <i>cas3</i> , <i>cas8e</i> , <i>cse2gr11</i> , <i>cas7</i> , 3 (12) ( <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , <i>csa3</i> , <i>cas5</i> , <i>cas6e</i> , <i>cas1</i> , <i>cas2</i> , and <i>csa3</i> ))
No. of prophages	10	16	12	18	14
Plasmids	1 [IncFIB(AP001918)]	3 [IncFIA, IncFIB(AP001918), IncX1]	0	3 [IncFIA, IncFIB(AP001918), IncX1]	2 [IncFIB(AP001918), IncY]
Predicted Antibiotic Resistance Genes (ARGs) in CARD	53	68	55	68	56
Predicted Virulence Factor Genes (VFGs) in VFDB	64	62	32	59	82
Predicted siderophores	Saccharide and enterobactin	Saccharide and enterobactin	Saccharide and enterobactin	Saccharide and enterobactin	Saccharide and enterobactin
Predicted bacteriocins	Bottomrymycin	Saccharides nd bottomrymycin	Bottomrymycin	Bottomrymycin	Bottomrymycin, 77.3;lin_M18
No. of subsystems (coverage, genes)	378 (33%, 2,112)	374 (32%, 2,079)	365 (34%, 1,981)	374 (32%, 2,082)	79 (31%, 2,086)

index by PathogenFinder v.1.1 (12), CRISPR arrays and prophages by CRISPRimmunity (13), plasmids by PlasmidFinder v.2.1 (14), antibiotic resistance genes (ARGs) by CARD v.3.2.4 with RGI main [the ARGs selection criteria were set to perfect (100% identity) and strict (>95% identity) hits only to the curated reference sequences in the CARD databases] (15) and ResFinder v.4.1 (16); virulence factor genes (VFGs) by VFDB with VFAnalyzer (17) and VirulenceFinder v.2.0 (18); siderophores by antiSMASH v.7.0 (19); bacteriocins by BAGEL suite v.4.0 (20); and metabolic functional features by RAST v.2.0 (21). Default parameters were used for all software unless otherwise specified.

Genomic characteristics of all the five *E. coli* strains are presented in Table 1. The *E. coli* strains MTR\_DSS\_V097, MTR\_DSS\_V098, MTR\_DNS\_V20, MTR\_DSR\_V504, and MTR\_DNS\_V06 contained 53, 68, 55, 68, and 56 predicted ARGs, respectively, and 64, 62, 32, 59, and 82 VFGs, respectively. Plasmids, such as IncFIA, IncFIB(AP001918), IncX1, and IncY were harbored in these strains. Four different sequence types (ST2, ST398, ST441, and ST1162) were identified in these strains, and the pathogenicity indices varied from 93% to 95%. The number of CRISPR arrays varied from two to three with 8–12 genes; the prophage numbers ranged between 10 and 18. Two types of siderophores and three types of bacteriocins were identified in the sequenced genomes. Moreover, 378, 374, 365, 359, and 79 subsystems were identified in the *E. coli* strains MTR\_DSS\_V097, MTR\_DSS\_V098, MTR\_DNS\_V20, MTR\_DSR\_V504, and MTR\_DNS\_V06, respectively.

## ACKNOWLEDGMENTS

This work was conducted as part of the CGIAR Research Initiative on Resilient Cities Through Sustainable Urban and Peri-urban Agri-food Systems and is supported by contributors to the CGIAR Trust Fund (<https://www.cgiar.org/funders>).

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Microbiology and Hygiene, Faculty of Veterinary Science, Bangladesh Agricultural University, Mymensingh, Bangladesh

<sup>2</sup>Department of Animal Sciences, University of California—Davis, Davis, California, USA

<sup>3</sup>World Vegetable Center, Tainan, Taiwan

<sup>4</sup>World Vegetable Center, Bangkok, Thailand

## AUTHOR ORCIDs

Md. Saiful Islam  <http://orcid.org/0000-0002-6870-4595>

Pritom Kumar Pramanik  <http://orcid.org/0000-0002-2862-8794>

Md. Liton Rana  <http://orcid.org/0000-0003-1400-1250>

Srinivasan Ramasamy  <http://orcid.org/0000-0002-8454-7331>

Pepijn Schreinemachers  <http://orcid.org/0000-0003-1596-3179>

Ricardo Oliva  <http://orcid.org/0000-0001-9517-5249>

Md. Tanvir Rahman  <http://orcid.org/0000-0001-5432-480X>

## FUNDING

Funder	Grant(s)	Author(s)
Consortium of International Agricultural Research Centers (CGIAR)		Md. Tanvir Rahman

## AUTHOR CONTRIBUTIONS

Md. Saiful Islam, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review and editing | Pritom Kumar Pramanik, Data curation, Investigation, Methodology | Md. Liton Rana, Investigation, Methodology | Srinivasan Ramasamy, Writing – review and editing | Pepijn Schreinemachers, Writing – review and editing | Ricardo Oliva, Supervision, Writing – review and editing

| Md. Tanvir Rahman, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review and editing

## DATA AVAILABILITY

See Table 1 for a list of accession numbers reported on.

## REFERENCES

1. Ballah FM, Islam MS, Levy S, Ferdous FB, Sobur MA, Rahman AT, Rahman M, Hoque MN, Hassan J, Rahman MT. 2023. Draft genome sequence of biofilm-forming methicillin-resistant *Staphylococcus aureus* MTR\_V1 strain isolated from a ready-to-eat food in Bangladesh. *Microbiol Resour Announc* 12:e0059723. <https://doi.org/10.1128/MRA.00597-23>
2. Islam MS, Rahman A, Hassan J, Rahman MT. 2023. Extended-spectrum beta-Lactamase in *Escherichia coli* isolated from humans, animals, and environments in Bangladesh: a one health perspective systematic review and meta-analysis. *One Health* 16:100526. <https://doi.org/10.1016/j.onehlt.2023.100526>
3. Waturangi DE, Hudiono F, Aliwarga E. 2019. Prevalence of pathogenic *Escherichia coli* from salad vegetable and fruits sold in Jakarta. *BMC Res Notes* 12:247. <https://doi.org/10.1186/s13104-019-4284-2>
4. Ratshilingano MT, du Plessis EM, Duvenage S, Korsten L. 2022. Characterization of multidrug-resistant *Escherichia coli* isolated from two commercial lettuce and spinach supply chains. *J Food Prot* 85:122–132. <https://doi.org/10.4315/JFP-21-125>
5. Hitchins AD, Feng P, Watkins WD, Rippey SR, Chandler LA. 1998. *Escherichia coli* and the coliform bacteria, p 4. In Food US, Drug Administration (ed), *Bacteriological Analytical manual*, 8th ed. AOAC International, Gaithersburg.
6. van Veen SQ, Claas ECJ, Kuijper EJ. 2010. High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. *J Clin Microbiol* 48:900–907. <https://doi.org/10.1128/JCM.02071-09>
7. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
9. Andrews S. 2010. Online. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
10. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. Refseq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
11. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50:1355–1361. <https://doi.org/10.1128/JCM.06094-11>
12. Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O. 2013. Pathogenfinder-distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 8:e77302. <https://doi.org/10.1371/journal.pone.0077302>
13. Zhou F, Yu X, Gan R, Ren K, Chen C, Ren C, Cui M, Liu Y, Gao Y, Wang S, Yin M, Huang T, Huang Z, Zhang F. 2023. Crisprimmunity: an interactive web server for CRISPR-associated important molecular events and modulators used in geNome editing Tool identifying. *Nucleic Acids Res* 51:W93–W107. <https://doi.org/10.1093/nar/gkad425>
14. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>
15. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-L, Cheng AA, Liu S, et al. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>
16. Florensa AF, Kaas RS, Clausen P, Aytan-Aktug D, Aarestrup FM. 2022. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* 8:000748. <https://doi.org/10.1099/mgen.0.000748>
17. Liu B, Zheng D, Zhou S, Chen L, Yang J. 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 50:D912–D917. <https://doi.org/10.1093/nar/gkab1107>
18. Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and vruencefinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* 4:e27943. <https://doi.org/10.4161/bact.27943>
19. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJM, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* 51:W46–W50. <https://doi.org/10.1093/nar/gkad344>
20. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. 2018. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res* 46:W278–W281. <https://doi.org/10.1093/nar/gky383>
21. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>