

UC Davis

UC Davis Previously Published Works

Title

Draft genome sequence of Escherichia coli MTR_GS_S1457 strain isolated from a soil sample of a vegetable garden in Gazipur, Bangladesh.

Permalink

<https://escholarship.org/uc/item/145340xf>

Journal

Microbiology Resource Announcements, 13(6)

Authors

Islam, Md Saiful

Pramanik, Pritom

Rana, Md

et al.

Publication Date

2024-06-11

DOI

10.1128/mra.00021-24

Peer reviewed

Draft genome sequence of *Escherichia coli* MTR_GS_S1457 strain isolated from a soil sample of a vegetable garden in Gazipur, Bangladesh

Md. Saiful Islam,^{1,2} Pritom Kumar Pramanik,¹ Md. Liton Rana,¹ Md. Ashek Ullah,¹ Fahim Haque Neloy,¹ Srinivasan Ramasamy,³ Pepijn Schreinemachers,⁴ Ricardo Oliva,³ Md. Tanvir Rahman¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We announce the sequence of the *Escherichia coli* MTR_GS_S1457 strain isolated from a soil sample of a vegetable gardening system for the first time in Bangladesh. With a length of 4,918,647 bp, this strain contained one plasmid, two CRISPR arrays, 54 predicted antibiotic resistance genes, and 81 predicted virulence factor genes.

KEYWORDS whole-genome sequencing, *Escherichia coli*, soil, environment, vegetable gardening system, Bangladesh

Food safety hinges on bacteria transmission from contaminated soil to edible plant parts, including *Escherichia coli* strains (1). While not all strains harm humans, enteropathogenic or enterohemorrhagic strains pose health risks (2). Manure used as fertilizer in family gardens could harbor such strains and potentially be transferred to humans (3). This first study in Bangladesh analyzes genome-based *E. coli* from soil in a family vegetable garden.

In November 2022, 2 grams of soil was aseptically collected from a family vegetable garden in the Gazipur district of Bangladesh (25.6135°N, 83.5070°E) by penetrating the soil surface (0–5 cm depth, 2.5 cm diameter). The soil was then homogenized and sieved using a 6.35 mm mesh sieve. The sample was then introduced into the nutrient broth (HiMedia, India; catalog number: M002-500G), incubated at 37°C for 24 hours, spread onto an eosin methylene blue agar plate, and the plate was then incubated overnight at 37°C. The colonies, showing metallic green sheens on the plate, underwent Gram staining and biochemical tests (indole and oxidase) to isolate *E. coli* (4). The *E. coli* MTR_GS_S1457 strain was cultured in nutrient broth aerobically overnight at 37°C, and the DNA extraction from the cultured broth was done using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Subsequently, the DNA library was prepared with the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). Genome sequencing was conducted on the Illumina NextSeq2000 platform, generating paired-end reads with a 2 × 150 bp length. The genome assembly was performed by Unicycler v0.4.9 (5), followed by the trimming of raw paired-end reads ($n = 2,902,088$) using Trimmomatic v0.39 (6) to remove Illumina adapters, known Illumina artifacts, and phiX reads from the data set. Quality assessment was performed through FastQC v0.11.7 (7). Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v6.6) (8). Sequence type was predicted using MLST v.2.0 (9) using the Center for Genomic Epidemiology database (<https://genomic epidemiology.org/services/>); serotype by SerotypeFinder v.2.0 (10); the pathogenicity index by PathogenFinder v.1.1 (11); CRISPR arrays and prophages by CRISPRimmunity (12); plasmids by PlasmidFinder v.2.1 (13); antibiotic resistance genes (ARGs) by the Comprehensive Antibiotic Resistance Database (CARD v.3.2.4) with RGI main v.6.0.2 (14); and virulence factor genes (VFGs)

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

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

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 28 January 2024

Accepted 29 April 2024

Published 17 May 2024

Copyright © 2024 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/).

by the virulence factor database with VFAnalyzer (15). Unless otherwise noted, default parameters were used for all software.

The assembled *E. coli* MTR_GS_S1457 strain comprised 112 contigs, with a guanine-cytosine content of 50.6%. The overall genome size was 4,918,647 bp with a coverage of 69.01x. The N_{50} for the assembled contigs was 222,202 bp. This genome had 4,832 genes, 4,749 protein-coding sequences, 83 RNA genes, 2 CRISPR arrays with nine signature genes (*cas2*, *cas1*, *cas6e*, *cas5*, *cas7*, *cse2gr11*, *cas8e*, *cas3*, and *WYL*), 13 prophages, and 1 plasmid. The serotype of the strain was predicted to be O51:H40, a probable enteropathogenic *E. coli*. With a pathogenicity index of 0.938, our genome was typed as ST19, based on eight-gene core-genome multilocus sequence typing (e.g., *dinB*, *icdA*, *pabB*, *polB*, *putP*, *trpA*, *trpB*, and *uidA*). Furthermore, the genome contained 54 predicted ARGs and 81 predicted VFGs.

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

¹Department of Microbiology and Hygiene, Faculty of Veterinary Science, Bangladesh Agricultural University, Mymensingh, Bangladesh

²Department of Animal Sciences, University of California—Davis, Davis, California, USA

³Safe and Sustainable Value Chains Flagship Program, World Vegetable Center, Shanhua, Tainan, Taiwan

⁴Healthy Diets and Enabling Impact Flagship Programs, World Vegetable Center—East and Southeast Asia, Bangkok, Thailand

AUTHOR ORCIDs

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

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

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

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 |

DATA AVAILABILITY

The WGS shotgun analysis of *E. coli* MTR_GS_V1777 was deposited to GenBank under the accession number [JAVTVQ000000000](https://www.ncbi.nlm.nih.gov/nuccore/JAVTVQ000000000). The relevant data, including the raw reads, were submitted with BioProject accession number [PRJNA1020214](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1020214), BioSample accession number [SAMN37518441](https://www.ncbi.nlm.nih.gov/biosample/SAMN37518441), and SRA accession number [SRR26151865](https://www.ncbi.nlm.nih.gov/sra/SRR26151865). In this version, the specific version being referred to is identified as [JAVTVQ000000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JAVTVQ000000000.1).

REFERENCES

- Ibekwe AM, Grieve CM, Yang CH. 2007. Survival of *Escherichia coli* O157:H7 in soil and on lettuce after soil fumigation. *Can J Microbiol* 53:623–635. <https://doi.org/10.1139/W07-003>
- Navab-Daneshmand T, Friedrich MND, Gächter M, Montealegre MC, Mlambo LS, Nhiwatiwa T, Mosler H-J, Julian TR. 2018. *Escherichia coli* contamination across multiple environmental compartments (soil, hands, drinking water, and Handwashing water) in urban Harare: correlations and risk factors. *Am J Trop Med Hyg* 98:803–813. <https://doi.org/10.4269/ajtmh.17-0521>
- Johannessen GS, Bengtsson GB, Heier BT, Bredholt S, Wasteson Y, Rørvik LM. 2005. Potential uptake of *Escherichia coli* O157:H7 from organic manure into crisphead lettuce. *Appl Environ Microbiol* 71:2221–2225. <https://doi.org/10.1128/AEM.71.5.2221-2225.2005>
- Hitchins AD, Feng P, Watkins WD, Rippey SR, Chandler LA, U.S. Food and Drug Administration. 1998. *Escherichia coli* and the Coliform bacteria, p

4. In *Bacteriological analytical manual*, 8th ed. AOAC International, Gaithersburg.
5. 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>
6. 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>
7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
8. 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>
9. 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>
10. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy *in Silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 53:2410–2426. <https://doi.org/10.1128/JCM.00008-15>
11. 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>
12. 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. Crisprimunity: an interactive web server for CRISPR-associated important molecular events and modulators used in genome identifi tool identifying. *Nucleic Acids Res* 51:W93–W107. <https://doi.org/10.1093/nar/gkad425>
13. 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>
14. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Boucharad M, Edalatmand A, Huynh W, Nguyen A-LV, 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>
15. 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>