Deciphering the complete genome sequence of multidrug-resistant *Escherichia coli* strain Hakim RU_GHWS isolated from ladies' hall sewage water in Bangladesh

Md. Arif-Uz-Zaman Polash,[1] Nusrat Zahan,[1] Muhib Ullah Khan,[1] Subir Sarker,[2,3] Tianyu Zhang,[4] Md. Hakimul Haque[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT This report details the genome sequence of *Escherichia coli* strain Hakim RU_GHWS, isolated from sewage water. The assembled genome comprises 5.022 Mb with 77.675× coverage, depicting an average GC content of 50.50%. This genome contains 10 CRISPR arrays, 14 prophages, 65 antibiotic resistance genes, and 28 virulence factor genes.

*E*scherichia coli, commonly found in the lower intestines of humans and animals, include strains that pose significant health risks (1). Sewage water increases antimicrobial resistance (AMR) due to dense bacterial environments and antibiotics, promoting horizontal gene transfer among bacteria (2). Contaminated water can spread resistant bacteria to drinking water, agriculture, and ecosystems, increasing human exposure (2). Multidrug-resistant *E. coli* in animals, humans, food, and environments highlight the need for AMR surveillance through the One Health approach to guide targeted public health interventions (3–6).

The Institute of Biological Sciences (IBScs) at the University of Rajshahi, Bangladesh, approved this study under Memo No. 56/321/IAMEBBC/IBScs. In September 2023, following standard procedures, we collected sewage water samples at the University of Rajshahi (24.3733°N, 88.6049°E). The samples were mixed thoroughly, transferred to sterile tubes, and transported to the laboratory. We then inoculated these samples on UTI agar (HiMedia, India) and incubated them aerobically at 37°C for 18–24 hours (6). *E. coli* was isolated using MacConkey (HiMedia, India) and eosin methylene blue agar (HiMedia, India), followed by staining and biochemical tests (7). Antibiogram study of the isolates was performed using the disk diffusion method (8), following CLSI guidelines (9). The strain presented resistance to gentamicin, amoxicillin, doxycycline, cephradine, ciprofloxacin, co-trimoxazole, and azithromycin. We cultured the isolated strain in nutrient broth (HiMedia, India) overnight at 37°C and then extracted its genomic DNA using the Qiagen DNA Mini Kit (QIAGEN, Hilden, Germany). The genomic DNA was enzymatically fragmented using the NEBNext dsDNA Fragmentase Kit (NEB, MA, USA), and size selection was carried out with SPRI beads (10). A sequencing library was prepared using the Nextera DNA Flex Library Preparation Kit (Illumina, San Diego, CA, USA), and the library was sequenced with 2×150 paired-end reads on the Illumina NextSeq2000 platform. Quality checks were performed using FastQC v0.11.7 (11). Raw reads ($n = 4,517,090$) were trimmed using Trimmomatic v0.39 (12), and genome assembly was conducted using Unicycler v0.4.9 (13). The annotation of the genome was carried out using PGAP v3.0 (14). The assembled genome was analyzed for antibiotic resistance genes (ARGs) using CARD v.3.2.4 with RGI v6.0.2 (15) and ResFinder v.4.1 (16),

TABLE 1 Genomic elements of the *E. coli* strain Hakim RU_GHWS

Elements	Values
Genome size	5,022,442 bp
Genome coverage	77.675×
G + C content	50.50%
Number of contigs	271
Contig L50	28
Contig N50	55,646 bp
Total genes	4,976
Coding sequences	4,918
Coding genes	4,733
RNA genes	58
tRNAs genes	48
rRNAs genes	2
tmRNAs gene	1
ncRNAs genes	8
Pseudo genes	185
Genes assigned to SEED subsystems	5,083
Number of subsystems	373

virulence factor genes using VFDB with VFanalyzer v.4.0 (17), pathogenicity index using PathogenFinder v.1.1 (18), sequence type using MLST v.2.0 (19), CRISPR arrays using CRISPRimmunity (20), and metabolic functional features using RAST v.2.0 (21). We used default parameters for all tools unless noted otherwise.

The features of the draft genomes are chronicled in Table 1. Notably, 65 ARGs, 28 virulence genes, and 6 plasmids [IncFIA(HI1), IncFIB(pB171), IncFII, IncHI1A, IncHI1B(R27), and IncX1] were predicted. MLST classified the genome as sequence type unknown but nearest to 999, and the PathogenFinder tool specified a pathogenicity index of 0.913. The genome exhibited 10 CRISPR arrays with 16 signature genes (*cas3, csa3, cas2, cas1, cas6e, cas5, cas7, cse2gr11, cas8e, TnsC, DinG, DEDDh, WYL, PD-DExK, c2c9_V-U4,* and *cas3HD*) and 14 prophages. RAST analysis uncovered 373 subsystems comprising 5,083 genes with 30% coverage (Fig. 1).
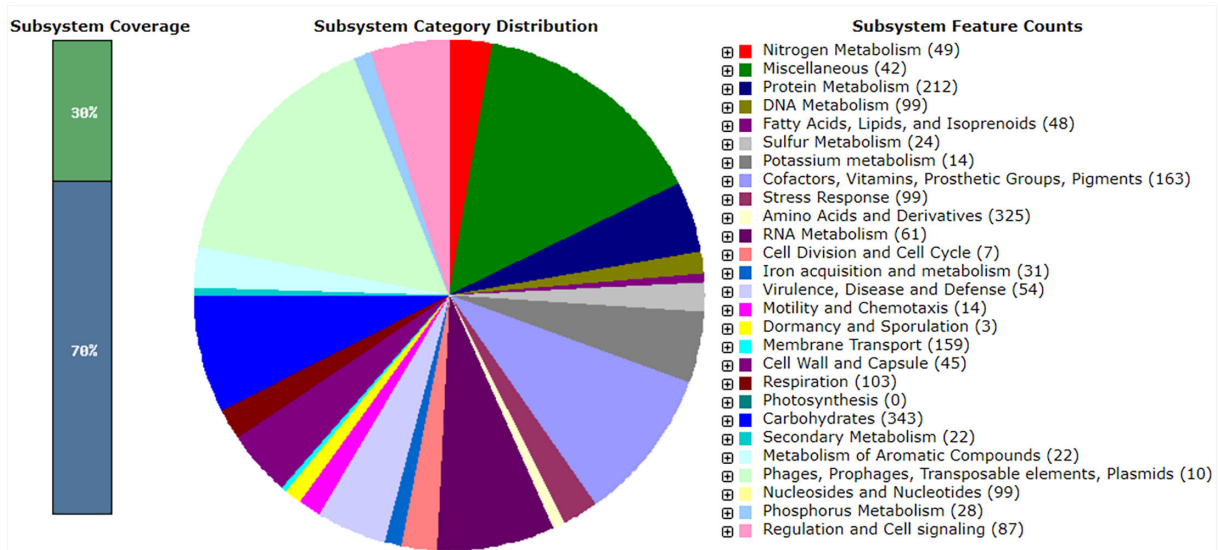

FIG 1 Metabolic functional features in the assembled genome of the *E. coli* Hakim RU_GHWS in SEED viewer. The 30% coverage indicates the completeness of functional roles within a specific subsystem across different genomes.

AUTHOR AFFILIATIONS

[1]Department of Veterinary and Animal Sciences, University of Rajshahi, Rajshahi, Bangladesh

[2]Biomedical Sciences & Molecular Biology, College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia

[3]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Queensland, Australia

[4]State Key Laboratory of Respiratory Disease, Guangzhou Institutes of Biomedicine and Health (GIBH), Chinese Academy of Sciences (CAS), Guangzhou, China

AUTHOR ORCIDs

Tianyu Zhang 🆔 http://orcid.org/0000-0001-5647-6014
Md. Hakimul Haque 🆔 http://orcid.org/0000-0002-3856-6478

AUTHOR CONTRIBUTIONS

Md. Arif-Uz-Zaman Polash, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft | Nusrat Zahan, Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft | Muhib Ullah Khan, Conceptualization, Data curation, Investigation, Methodology | Subir Sarker, Software, Writing – review and editing | Tianyu Zhang, Software, Writing – review and editing | Md. Hakimul Haque, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The study on E. coli strain Hakim RU_GHWS, conducted using the WGS shotgun approach, was submitted to NCBI/GenBank, and the assembly was deposited under the accession number JBEHGP000000000. The pertinent data, including the original readings, were stored with BioProject accession number PRJNA1101920, BioSample accession number SAMN41006273, and SRA accession number SRR28733676. The specific version mentioned in this document is labeled as JBEHGP000000000.1.

REFERENCES

1. Navab-Daneshmand T, Friedrich MND, Gächter M, Montealegre MC, Mlambo LS, Nhiwatiwa T, Mosler HJ, 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
2. Karkman A, Do TT, Walsh F, Virta MPJ. 2018. Antibiotic-resistance genes in waste water. Trends Microbiol 26:220–228. https://doi.org/10.1016/j.tim.2017.09.005
3. Bakshi P, Bhowmik A, Ahsan S, Alim SR. 2024. Identification of antibiotic-resistant pathogens and virulence genes in *Escherichia coli* isolates from food samples in the Dhaka University campus of Bangladesh. Food Sci Nutr 12:1995–2002. https://doi.org/10.1002/fsn3.3896
4. Zinnah MA, Bari MR, Islam MT, Hossain MT, Rahman MT, Haque MH, Babu SAM, Ruma RP, Islam MA. 2007. Characterization of *Escherichia coli* isolated from samples of different biological and environmental sources. Bangl J Vet Med 5:25–32. https://doi.org/10.3329/bjvm.v5i1.1305
5. Islam MS, Rahman AMMT, 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
6. Haque MH, Islam MZ, Miah ML, Das SK, Sarker S. 2023. Molecular characterization of multidrug-resistant *Escherichia coli* isolated from human urine infections with their antibiogram profile. J Adv Biotechnol Exp Ther 6:172–183. https://doi.org/10.5455/jabet.2023.d116
7. Bergey DH, Buchanan RE, Gibbons NE. 1974. Bergey's manual of determinative bacteriology. 8th ed, p 966–1097. American Society for Microbiology, Washington, DC, USA.
8. Bauer AW, Kirby WMM, Sherris JC, Turck M. 1966. Antibiotic susceptibility testing by a standardized single disk method. Am J Clin Pathol 45:493–496. https://doi.org/10.1093/ajcp/45.4_ts.493
9. 2022. M100-S32 performance standards for antimicrobial susceptibility testing. Clinical and Laboratory Standards Institute, Wayne, PA, USA.
10. Liu D, Li Q, Luo J, Huang Q, Zhang Y. 2023. An SPRI beads-based DNA purification strategy for flexibility and cost-effectiveness. BMC Genomics 24:125. https://doi.org/10.1186/s12864-023-09211-w
11. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data

12. 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

13. 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

14. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

15. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard 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

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. 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

19. 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

20. 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

21. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75