Microbiology Resource Announcements

| Genetics and Molecular Biology | Announcement

Whole-genome sequencing of *Escherichia coli* DH5α using MinION nanopore technology

Megha S. Kumar,[1,2] P.V. Suneesh,[1,2] Manoj Bhat Krishna,[3] K.P. Soman,[3] T.G. Satheesh Babu[1,2,4]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We present the whole-genome sequence of *Escherichia coli* DH5α, a laboratory strain integral to molecular biology applications. We assembled a high-quality genome using Qiagen's DNeasy Blood and Tissue Kit and the Oxford Nanopore MinION platform (~4.58 Mb, 50.8% guanine-cytosine [GC] content), offering insights into genetic structure and functional characteristics.

KEYWORDS whole-genome sequencing, *E. coli*, *de novo* assembly, MinION

Escherichia coli is a gram-negative bacteria commonly associated with warm-blooded animals and a core member of the *Enterobacteriaceae* family (1). The species exhibits considerable genomic and phenotypic diversities shaped by adaptation to distinct ecological niches, resulting in strain-specific metabolic capabilities and stress responses (2, 3). This makes *E. coli* an excellent model organism for investigating bacterial physiology, genetics, and evolution.

E. coli DH5α is a widely used laboratory strain derived from the K-12 lineage and engineered for high-efficiency cloning applications (4). Its key genetic features include the lacZΔM15 mutation enabling blue-white screening, the recA1 and endA1 mutations that improve insert stability and plasmid DNA quality, and the hsdR17 allele, which permits efficient transformation of methylated DNA (4). A closed genome sequence has previously been reported for NEB 5-alpha, a commercial *fhuA2* derivative of DH5α that serves as a finished reference for this lineage (5). However, NEB 5-alpha represents a proprietary laboratory derivative rather than a culture-collection DH5α isolate. To provide an independent long-read reference genome for a widely distributed DH5α strain, this study sequenced and assembled *E. coli* DH5α obtained from the Microbial Type Culture Collection (MTCC 1652), Chandigarh, India.

Cells were cultured in Luria-Bertani broth at 37°C with shaking at 200 rpm for 10 h before DNA extraction. Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen) following the protocol for gram-negative bacteria. No deliberate DNA fragmentation step was performed; high-molecular-weight DNA was preserved throughout extraction and library preparation, with only minimal mechanical stress introduced during routine pipetting and clean-up steps. Library preparation was performed with the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies [ONT]), and sequencing was conducted on the MinION platform with a FLO-MIN106 flow cell (R9.4.1). AMPure XP bead clean-up was performed using short fragment buffer to retain fragments across the full size distribution without additional size selection. Read acquisition and quality were monitored using MinKNOW, and high-accuracy basecalling with adapter trimming was performed using Guppy v3.4.4 with the dna_r9.4.1_450bps_hac.cfg configuration.

Sequencing produced 81,605 reads with a read length N50 of 10,380 bp, a mean read length of 5,974 bp, a mean Phred quality score of 11.1, and an estimated genome coverage of 93×, as assessed using NanoPlot (6). *De novo* assembly was performed

Editor Catherine Putonti, Loyola University Chicago, Chicago, Illinois, USA

Address correspondence to T.G. Satheesh Babu, tg_satheesh@cb.amrita.edu.

The authors declare no conflict of interest.

Received 23 September 2025
Accepted 22 February 2026
Published 12 March 2026

with Unicycler v0.5.0 (7) in long-read-only mode, followed by three rounds of polishing with Racon. The circular chromosome was automatically rotated by Unicycler during assembly to begin at the dnaA gene. Additional consensus polishing was performed using Medaka v1.12.1 (8) with the r941_min_high_g306 model. The genome annotation was performed with National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) v6.10 (9). All the bioinformatics tools used default parameters, unless otherwise stated.

The final assembly comprises a single circular chromosome of 4,583,112 bp with a guanine-cytosine (GC) content of 50.84% and no plasmids. PGAP identified 4,542 genes, including 4,421 protein-coding sequences and 121 RNA genes. Two CRISPR arrays were detected using CRISPRCasFinder v4.3.2 (10), with consensus repeat sequences GTGTT CCCCGCGCCAGCGGGGATAAACC and GAGTTCCCCGCGCCAGCGGGGATAAACC, which are consistent with conserved type I-E CRISPR repeats in *E. coli* (11).

ACKNOWLEDGMENTS

This work is supported by the TCS Research Scholarship Program (Corporate Research and Innovation) and Amrita Vishwa Vidyapeetham, Coimbatore.

AUTHOR AFFILIATIONS

[1]Department of Chemistry, Amrita School of Physical Sciences, Amrita Vishwa Vidyapeetham, Coimbatore, India
[2]Amrita Biosensor Research Lab, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
[3]Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India
[4]Amrita Biomedical Engineering Centre, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

AUTHOR ORCIDs

Megha S. Kumar http://orcid.org/0000-0003-4278-8361
T.G. Satheesh Babu http://orcid.org/0000-0002-9919-4372

AUTHOR CONTRIBUTIONS

Megha S. Kumar, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft | P.V. Suneesh, Writing – review and editing | Manoj Bhat Krishna, Supervision, Writing – review and editing | K.P. Soman, Writing – review and editing.

DATA AVAILABILITY

The raw reads are available under accession number SRR31302084, and the assembled genome is available under accession number CP199401.1 in NCBI under BioProject accession number PRJNA1184340.

REFERENCES

1. Ramos S, Silva V, Dapkevicius M de LE, Caniça M, Tejedor-Junco MT, Igrejas G, Poeta P. 2020. *Escherichia coli* as commensal and pathogenic bacteria among food-producing animals: health implications of extended spectrum β-lactamase (ESBL) production. Animals (Basel) 10:2239. https://doi.org/10.3390/ani10122239
2. NandaKafle G, Huegen T, Potgieter SC, Steenkamp E, Venter SN, Brözel VS. 2021. Niche Preference of *Escherichia coli* in a Peri-Urban pond ecosystem. Life (Basel) 11:1020. https://doi.org/10.3390/life11101020
3. Koh XP, Shen Z, Woo CF, Yu Y, Lun HI, Cheung SW, Kwan JKC, Lau SCK. 2022. Genetic and ecological diversity of *Escherichia coli* and cryptic *Escherichia* Clades in subtropical aquatic environments. Front Microbiol 13:811755. https://doi.org/10.3389/fmicb.2022.811755
4. Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580. https://doi.org/10.1016/s0022-2836(83)80284-8
5. Anton BP, Raleigh EA. 2016. Complete genome sequence of NEB 5-alpha, a derivative of *Escherichia coli* K-12 DH5α. Genome Announc 4:e01245-16. https://doi.org/10.1128/genomeA.01245-16
6. De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39:btad311. https://doi.org/10.1093/bioinformatics/btad311
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. Oxford Nanopore Technologies. 2026. Nanoporetech/medaka (v1.12.1). Oxford Nanopore Technologies. https://github.com/nanoporetech/medaka.
9. 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
10. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCas-Finder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246–W251. https://doi.org/10.1093/nar/gky425
11. Xue C, Sashital DG. 2019. Mechanisms of type I-E and I-F CRISPR-cas systems in *Enterobacteriaceae* EcoSal Plus 8. https://doi.org/10.1128/ecosalplus.ESP-0008-2018