AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Environmental Microbiology | Announcement

Draft genome sequence of heat-tolerant *Escherichia coli* strain AKS1GF CIFRI isolated from a coal power plant discharge in the river ganga

Amiya Kumar Sahoo,[1] Dharmendra Kumar Meena,[1] Basanta Kumar Das,[1] Subhankhi Dasgupta,[1] Smruti Samantaray,[1] Debasmita Mohanty,[1] Ayushman Gadnayak,[1] Subhasmita Behera,[1] Asit Kumar Bera[1]

AUTHOR AFFILIATION See affiliation list on p. 2.

ABSTRACT *Escherichia coli* strain, AKS1GF ICAR-CIFRI, was isolated from a thermally contaminated water with average water temperatures of 45°C ± 2.5°C in river Ganga, India. The draft genome sequence is of 4.7 MB consisting of 44 contigs. The strain contains 4,314 protein coding genes and 96 RNA genes.

KEYWORDS *Escherichia coli*, genome sequence, thermal discharge, river ganga

Escherichia coli, a widely distributed bacteria often found in the gut microbiota, has been shown to inhabit several ecological habitats, including aquatic environment (1). In order to examine the genetic foundation of *E. coli's* ability to adapt at high-temperature habitat, we conducted a genome sequencing analysis of AKS1GF that was obtained from a thermal discharge site in the river Ganga, the largest river of India with cultural and ecological significance (2).

The *E. coli* strain AKS1GF was isolated from a soil sample collected from the thermal discharge site of a coal power plant in the river Ganga during April 2024 (87°53'39"E-87°56'41"E;24°40'19"N-24°52'12"N). The soil sample was serially diluted, and 10^{-5} dilution was transferred to Escherichia coli (EC) broth (Himedia, Mumbai, India) and incubated at 37°C for 24 hours. One loopfull of culture was transferred to MacConkey agar (Himedia, Mumbai, India) plate for isolation of pure culture. The single lactose-producing bacteria were picked from the MacConkey agar plate and transferred to Tryptone Soya Broth (TSB) for genomic DNA isolation. Genomic DNA was isolated using QIAamp DNA mini kit (Qiagen). A multiplex PCR was used to confirm the *E. coli* isolate by using five different primers (3), uidA (ß-D-galactosidase), lacZ (ß-D-galactosidase), lacY (lactose permease), cyd (cytochrome bd complex), and phoA (bacterial alkaline phosphatase).

Using the QIASeq FX DNA kit (Qiagen), a paired-end library containing the genomic DNA was created and used for sequencing on the Illumina NextSeq 2000 platform. A total of 5.5 Gb was generated (Table 1). To check the quality of the data, FASTQ files was pre-processed using FastQc v0.11.9 (4), following the specifications in MultiQC report ((5). To filter adapter contamination, the fastp tool (v0.12.4) (6) was used. The genome was assembled with Megahit (7). The genome was made up of 44 contigs and had a N50 value 239.8 kb, as indicated in Table 1. Annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (8) and revealed 4,501 genes, with 4,405 coding sequences and 96 noncoding sequences (12 rRNAs, 74 tRNAs, and 10 noncoding RNAs).

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Amiya Kumar Sahoo, amiya7@gmail.com.

The authors declare no conflict of interest.

Received 30 August 2024
Accepted 5 October 2024
Published 29 October 2024

TABLE 1 Overview of the draft whole-genome sequence of *E. coli* strain AKS1GF

Sequencing method	Illumina NextSeq 2000
No of reads (bp)	20,000,000 (paired-end)
Average length of the paired-end reads	119 bp
Genome size	4.7 mb
No. of contigs	44
N50 (bp)	239.8 kb
GC%	51
Genome coverage	50.7×
Protein coding genes	4,314
Completeness	99.52%
Contamination	0.44%
Biosample accession no	SAMN42748805

ACKNOWLEDGMENTS

The authors express gratitude to the Director, ICAR-CIFRI, for the guidance in carrying out the sequencing result analysis. The authors acknowledge Indian Network for Fisheries and Animal Antimicrobial Resistance (INFAAR) and National Mission for Clean Ganga (NMCG) for the support. Special thanks to Atul Kumar and Aritriya Jana, Research Staff, ICAR-CIFRI for collecting the samples.

AUTHOR AFFILIATION

[1]ICAR-Central Inland Fisheries Research Institute, Barrackpore, Kolkata, India

AUTHOR ORCIDs

Amiya Kumar Sahoo http://orcid.org/0000-0002-3183-8565
Basanta Kumar Das http://orcid.org/0000-0002-6629-8992

AUTHOR CONTRIBUTIONS

Amiya Kumar Sahoo, Conceptualization, Formal analysis, Funding acquisition, Investigation | Dharmendra Kumar Meena, Conceptualization, Formal analysis, Investigation, Methodology | Basanta Kumar Das, Funding acquisition, Supervision | Subhankhi Dasgupta, Methodology, Validation | Smruti Samantaray, Formal analysis | Debasmita Mohanty, Methodology | Ayushman Gadnayak, Data curation, Methodology, Validation, Writing – original draft | Subhasmita Behera, Data curation, Methodology, Writing – original draft, Writing – review and editing | Asit Kumar Bera, Investigation, Writing – review and editing

DATA AVAILABILITY

The whole-genome sequence has been published in the NCBI DDBJ/EMBL/GenBank database under the accession number JBGDMG000000000. The version mentioned in this study is JBGDMG000000000.1, while the raw reads were submitted to SRA under run number SRR29979039.

REFERENCES

1. Woodward SE, Krekhno Z, Finlay BB. 2019. Here, there, and everywhere: how pathogenic *Escherichia coli* sense and respond to gastrointestinal biogeography. Cell Microbiol 21. https://doi.org/10.1111/cmi.13107
2. Chaturvedi S, Chakraborty B, Min L, Kumar A, Pathak B, Kumar R, Yu ZG. 2024. Exploring the dynamic microbial tapestry of South Asian rivers: insights from the Ganges and Yamuna ecosystems. Ecohydrology 17. https://doi.org/10.1002/eco.2662
3. Modak R, Das Mitra S, Krishnamoorthy P, Bhat A, Banerjee A, Gowsica BR, Bhuvana M, Dhanikachalam V, Natesan K, Shome R, Shome BR, Kundu TK. 2012. Histone H3K14 and H4K8 hyperacetylation is associated with

Escherichia coli-induced mastitis in mice. Epigenetics 7:492–501. https://doi.org/10.4161/epi.19742

4. Andrews S. 2010. FASTQC. a quality control tool for high throughput sequence data
5. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354
6. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all in-one FASTQ preprocessor. Biorxiv. https://doi.org/10.1101/274100
7. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033
8. 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