AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Escherichia coli* Strain Tj, Isolated from the Varzob River in Tajikistan

Munavvara Dzhuraeva,[a,c] Mehrangez Shokirova,[a] Ani Azaryan,[b,c] Hovik Panosyan,[b] Khursheda Bobodzhanova,[a] Nils-Kåre Birkeland[c]

[a]Center of Biotechnology, Tajik National University, Dushanbe, Tajikistan
[b]Department of Biochemistry, Microbiology, and Biotechnology, Yerevan State University, Yerevan, Armenia
[c]Department of Biological Sciences, University of Bergen, Bergen, Norway

ABSTRACT The 4.6-Mbp draft genome sequence of *Escherichia coli* strain Tj, isolated from the Varzob River in Tajikistan, is presented. This strain possesses four prophage elements related to *Shigella* phage SfV, *E. coli* O157:H7-specific phage ϕV10, lambdoid phage HK225, and coliphage Ayreon. It contains a gene encoding a hemolysin E toxin.

Escherichia coli (Migula 1895) Castellani and Chalmers 1919 is a Gram-negative, facultatively anaerobic, enteric bacterium that inhabits the intestines of warm-blooded animals and humans and frequently contaminates environmental waters and food. *E. coli* strain Tj was isolated from the Varzob River in Tajikistan, during a laboratory course at the Center of Biotechnology of the Tajik National University, on Endo LES agar plates at 44°C as a dark-red colony and confirmed as *E. coli* by its API 20E profile (5-0-4-4-5-5-2). The 16S rRNA gene sequence, obtained by PCR amplification and Sanger sequencing as described elsewhere (1), shared up to 99.86% sequence identity with *E. coli* strains in BLASTn searches. For genome sequencing, the strain was cultivated in LB overnight at 37°C with shaking. DNA was extracted using the GenElute bacterial genomic DNA kit (Sigma-Aldrich). The genome was sequenced at Eurofins Genomics using a NEBNext Ultra II DNA library preparation kit and Illumina HiSeq 2500 paired-end sequencing technology with a read length of 2 × 150 bp, yielding 6,416,190 reads and 1,924,857,000 sequenced bases. Reads with a maximum of 7 bases with a Phred score below 28 were initially discarded, and additional quality control procedures were performed using the Trim Reads tool in the CLC Genomics Workbench v. 8.5.1. Unless otherwise stated, all software was used with default values. Assembly was performed using the CLC *de novo* assembly tool, resulting in 4,627,784 bp of unique sequence data distributed into 96 contigs with an N_{50} value of 111,402 bp, coverage of 417×, and GC content of 50.8%. The draft genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok). Genome completeness was estimated as 99.96% by CheckM v. 1.0.18 (2). A phylogenomic analysis, including a selection of *E. coli* strains and *Shigella* species, revealed clustering within the *E. coli/Shigella* clade, with pairwise average nucleotide identity (ANI) values of ≥97.0% (Fig. 1A). Only one enterobacterial toxin-encoding gene, a pore-forming hemolysin E homolog, was detected in strain Tj in a search of the 2016 release of the Virulence Factor Database (http://www.mgc.ac.cn/VFs) (3) using the VFanalyzer tool. However, many genomic islands (GIs), four prophage elements (P), and one CRISPR/Cas element were found (Fig. 1B) using IslandViewer 4 (4), Prophage Hunter (5), and CRISPRCasFinder (6), respectively. The prophage elements were related to *E. coli* phage Ayreon (P-1) (7), the serotype-converting temperate phage SfV from *Shigella flexneri* (P-2) (8), lambdoid coliphage HK225 (GenBank accession

Citation Dzhuraeva M, Shokirova M, Azaryan A, Panosyan H, Bobodzhanova K, Birkeland N-K. 2020. Draft genome sequence of *Escherichia coli* strain Tj, isolated from the Varzob River in Tajikistan. Microbiol Resour Announc 9:e00867-20. https://doi.org/10.1128/MRA.00867-20.

Editor David Rasko, University of Maryland School of Medicine

Address correspondence to Nils-Kåre Birkeland, nils.birkeland@uib.no.

Received 29 July 2020
Accepted 16 September 2020
Published 8 October 2020

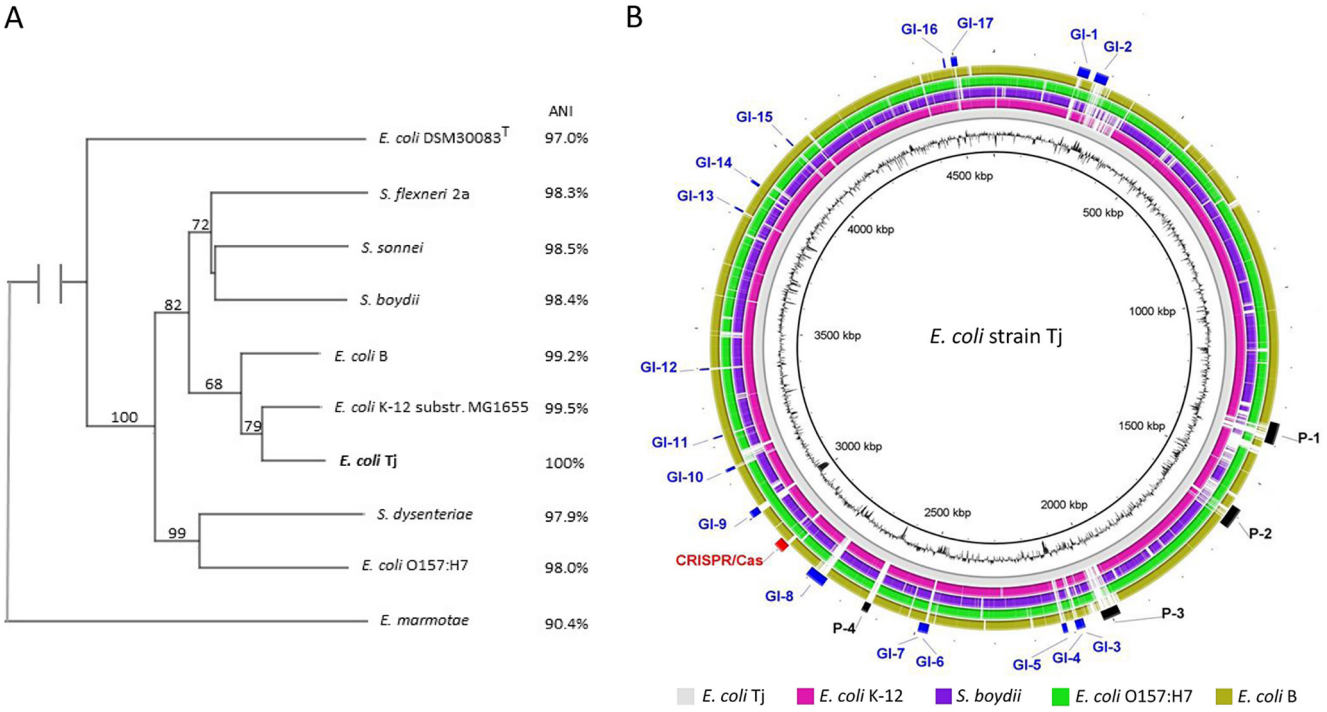

FIG 1 Genome-based phylogenetic affiliation of strain Tj with representative *E. coli* strains and *Shigella* spp., represented as a phylogenetic tree (A), and circular representation of the *E. coli* Tj genome as a reference, compared with representative *E. coli* strains and *Shigella* spp. (B), using the BLAST Ring Image Generator (BRIG) (10). The tree was inferred with FastME v. 2.1.6.1 (11) from Genome Blast Distance Phylogeny (GBDP) distances calculated from genome sequences using the TYGS server (https://tygs.dsmz.de) (12) and rooted with *Escherichia marmotae* as an outgroup. The branch lengths are scaled in terms of the GBDP distance formula d5. The numbers above the branches are GBDP pseudobootstrap support values of ≥68% from 100 replications. The tree was rooted at the midpoint (13). Before being used as a reference genome for the BRIG, the Tj contigs were ordered using Mauve (14) with the *E. coli* K-12 genome as the template and merged. Ring color codes are indicated below the rings. The black (innermost) ring indicates the GC content of strain Tj. Selected genomic features of strain Tj are indicated as GIs, prophage elements (P), and a CRISPR/Cas element. Genome sequence accession numbers are as follows: GCA_007049865.1 (*E. coli* Tj), NC_000913.3 (*E. coli* strain K-12 substrain MG1655), NC_002695.2 (*E. coli* O157:H7 strain Sakai), NZ_CP014268.2 (*E. coli* B), NZ_CP025979.1 (*E. marmotae*), NZ_CP026731.1 (*Shigella boydii*), NZ_CP026774.1 (*Shigella dysenteriae*), NZ_CP026788.1 (*Shigella flexneri* 2a), CP026802.1 (*Shigella sonnei*), and NZ_CP033092 (*E. coli* DSM 30083^T).

number NC_019717), and phage ϕV10, which specifically infects *E. coli* O157:H7 (P-4) (9). Most of the GIs encoded mobile elements and hypothetical genes. GI-1, GI-9, and GI-10 encoded type III and type VI secretion systems and a general secretion pathway, respectively. The reported data will be useful for future understanding of the genetic diversity and virulence potential of *E. coli* in Central Asia and the distribution, evolution, and dissemination of coliphages.

Data availability. The partial 16S rRNA gene and whole-genome shotgun sequences of *E. coli* strain Tj have been deposited in DDBJ/ENA/GenBank under accession numbers MT920313 and GCA_007049865.1, respectively. The associated BioProject, SRA, and BioSample accession numbers are PRJNA506592, SRR12336902, and SAMN10464432, respectively.

ACKNOWLEDGMENT

This work was funded by the Eurasia Program of the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (Diku) (grants CPEA-LT-2016/10095 and CPEA-LT-2017/10061).

REFERENCES

1. Shurigin V, Hakobyan A, Panosyan H, Egamberdieva D, Davranov K, Birkeland NK. 2019. A glimpse of the prokaryotic diversity of the Large Aral Sea reveals novel extremophilic bacterial and archaeal groups. MicrobiologyOpen 8:e00850. https://doi.org/10.1002/mbo3.850.
2. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
3. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative

pathogenomic platform with an interactive Web interface. Nucleic Acids Res 47:D687–D692. https://doi.org/10.1093/nar/gky1080.
4. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FSL. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45:W30–W35. https://doi.org/10.1093/nar/gkx343.
5. Song W, Sun HX, Zhang C, Cheng L, Peng Y, Deng Z, Wang D, Wang Y, Hu M, Liu W, Yang H, Shen Y, Li J, You L, Xiao M. 2019. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res 47:W74–W80. https://doi.org/10.1093/nar/gkz380.
6. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron 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.
7. Vlot M, Nobrega FL, Wong CFA, Liu Y, Brouns SJJ. 2018. Complete genome sequence of the *Escherichia coli* phage Ayreon. Genome Announc 6:e01354-17. https://doi.org/10.1128/genomeA.01354-17.
8. Allison GE, Angeles D, Tran-Dinh N, Verma NK. 2002. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri*. J Bacteriol 184:1974–1987. https://doi.org/10.1128/jb.184.7.1974-1987.2002.
9. Perry LL, SanMiguel P, Minocha U, Terekhov AI, Shroyer ML, Farris LA, Bright N, Reuhs BL, Applegate BM. 2009. Sequence analysis of *Escherichia coli* O157:H7 bacteriophage Phi V10 and identification of a phage-encoded immunity protein that modifies the O157 antigen. FEMS Microbiol Lett 292:182–186. https://doi.org/10.1111/j.1574-6968.2009.01511.x.
10. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402. https://doi.org/10.1186/1471-2164-12-402.
11. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol 32:2798–2800. https://doi.org/10.1093/molbev/msv150.
12. Meier-Kolthoff JP, Goker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3.
13. Farris JS. 1972. Estimating phylogenetic trees from distance matrices. Am Nat 106:645–668. https://doi.org/10.1086/282802.
14. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.