Complete Genome Sequence of Enterotoxigenic *Escherichia coli* Siphophage LL5

Denish Piya,[a,b]* Lauren Lessor,[b] Mei Liu,[b] Jason J. Gill[b,c]

[a]Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas, USA
[b]Center for Phage Technology, Texas A&M University, College Station, Texas, USA
[c]Department of Animal Science, Texas A&M University, College Station, Texas, USA

ABSTRACT Here, we describe the complete genome sequence of siphophage LL5. LL5 is a T1-like phage isolated against enterotoxigenic *Escherichia coli*, which causes traveler's diarrhea. LL5 is included as a component phage in the commercial prebiotic product PreforPro.

Enterotoxigenic *Escherichia coli* (ETEC) strains are characterized by the presence of heat-labile enterotoxins (LT) and/or heat-stable enterotoxins (HT) (1). ETEC strains can induce traveler's diarrhea (TD), resulting in watery diarrhea (1) which may be accompanied by nausea, vomiting, abdominal pain, fever, or blood in stool (2). TD typically self-resolves or may be successfully treated with antibiotics, but the global increase in the emergence of antibiotic resistance warrants the evaluation of alternative antibacterial approaches, such as the use of phages (3).

Phage LL5 was isolated by the enrichment method (4) from filter-sterilized (0.22-μm) influent from a municipal water treatment facility in College Station, TX, against a clinical ETEC isolate. Phage LL5 is also able to infect *E. coli* K-12 strains and was propagated on the *E. coli* strain DH5-alpha. The phage DNA was purified from high-titer lysates by a modified Wizard DNA purification kit (Promega) as previously described (5). Phage LL5 DNA was prepared using the GS FLX Titanium general DNA library preparation kit and was sequenced by FLX Titanium 454 pyrosequencing at the Emory-Georgia Research Alliance Genomics Center (Emory University, GA); trimmed sequence reads were assembled into a single contig at 19.9-fold coverage using Newbler version 2.5.3 (454 Life Sciences) at default settings. The contig was closed, based on its circular assembly, which produced identical sequences at each end. Structural annotation was conducted using GLIMMER version 3.0 (6) and MetaGeneAnnotator version 1.0 (7), with tRNAs predicted by ARAGORN version 2.36 (8) and verified by tRNAscan-SE (9) and gene functions predicted by InterProScan version 5.15-54.0 (10), the NCBI Conserved Domain Database (11), TMHMM version 2.0 (http://www.cbs.dtu.dk/services/TMHMM), BLASTp version 2.2.28 (12), and HHpred version 2.1 (13). Phage genome annotation was conducted using the Phage Galaxy (11) and Web Apollo (12) instances hosted by the Center for Phage Technology (https://cpt.tamu.edu). Phages were imaged by transmission electron microscopy at the Texas A&M University Microscopy and Imaging Center as previously described (14, 15).

LL5 is a siphophage with a head diameter of ~60 nm and a flexible tail ~150 nm long. LL5 has a genome of 49,788 bp with 42.5% G+C content, 88 predicted protein-coding genes, and no tRNA genes. Phage LL5 is closely related to the T1-like coliphage TLS (NCBI reference sequence NC_009540) (16), with 96% DNA sequence identity over 90% of the LL5 genome based on BLASTn. The genome produced a circular assembly and was reopened to be syntenic with other T1-like phages in the NCBI database, such as TLS (NC_009540) and T1 (NC_005833). Thirty-three LL5-encoded proteins could be

Address correspondence to Mei Liu, meiliu@tamu.edu, or Jason J. Gill, jason.gill@tamu.edu.

* Present address: Denish Piya, Department of Bioengineering, University of California, Berkeley, California, USA.

assigned putative functions. Genes identified included a DNA primase/helicase (gp58), ATP-dependent helicase (gp60), and helicase (gp75). Structural proteins identified included the portal protein (gp31), major capsid protein (gp36), minor tail proteins (gp41, gp42, gp47, and gp48), tail tube protein (gp43), tape measure protein (gp46), and tail fiber proteins (gp51 and gp57). The small and large terminase subunits were identified as gp29 and gp30, respectively. Like its T1-like relatives, LL5 encodes a canonical lysis cassette composed of a holin (gp70), endolysin (gp71), and unimolecular spanin (gp72). Phage LL5 is currently a component of the prebiotic product PreforPro.

Data availability. The annotated phage genome is deposited in NCBI GenBank under accession no. MH491968. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR9134807, and SAMN11874326, respectively.

REFERENCES

1. Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol 2:123–140. https://doi.org/10.1038/nrmicro818.
2. Taylor DN, Hamer DH, Shlim DR. 2017. Medications for the prevention and treatment of travellers' diarrhea. J Travel Med 24:S17–S22. https://doi.org/10.1093/jtm/taw097.
3. Tribble DR. 2017. Resistant pathogens as causes of traveller's diarrhea globally and impact(s) on treatment failure and recommendations. J Travel Med 24:S6–S12. https://doi.org/10.1093/jtm/taw090.
4. Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, Kane A, Keene J, Maddox M, Martin R, Mohan S, Thorn AM, Russell DH, Young R. 2012. The Caulobacter crescentus phage phiCbK: genomics of a canonical phage. BMC Genomics 13:542. https://doi.org/10.1186/1471-2164-13-542.
5. Summer EJ. 2009. Preparation of a phage DNA fragment library for whole genome shotgun sequencing. Methods Mol Biol 502:27–46. https://doi.org/10.1007/978-1-60327-565-1_4.
6. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641. https://doi.org/10.1093/nar/27.23.4636.
7. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387–396. https://doi.org/10.1093/dnares/dsn027.
8. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
9. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
10. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
11. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129.
12. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
13. Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408.
14. Valentine RC, Shapiro BM, Stadtman ER. 1968. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. Biochemistry 7:2143–2152. https://doi.org/10.1021/bi00846a017.
15. Piya D, Vara L, Russell WK, Young R, Gill JJ. 2017. The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. Mol Microbiol 105:399–412. https://doi.org/10.1111/mmi.13705.
16. German GJ, Misra R. 2001. The TolC protein of *Escherichia coli* serves as a cell-surface receptor for the newly characterized TLS bacteriophage. J Mol Biol 308:579–585. https://doi.org/10.1006/jmbi.2001.4578.