AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of Enterotoxigenic *Escherichia coli* Myophage LL12

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

[a]Department of Biochemistry and 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 This work describes the complete genome sequence of the virulent myophage LL12. The 136-kb LL12 genome is related to coliphage V5 and is a component in the prebiotic PreforPro. LL12 was isolated against enterotoxigenic *Escherichia coli*, which causes traveler's diarrhea.

Enterotoxigenic *Escherichia coli* (ETEC) strains can be distinguished from other *E. coli* pathotypes by the presence of enterotoxins which induce traveler's diarrhea (TD), which is characterized by mild to severe watery diarrhea (1). Typically, TD self-resolves or may be treated with antibiotics, but due to the increase in the emergence of antibiotic resistance, alternative treatment approaches, such as phage therapy, have been suggested (2).

Phage LL12 was isolated from a municipal water treatment plant in College Station, TX, by growing a deidentified clinical ETEC isolate in the wastewater sample supplemented with LB broth (Difco) at 37°C with aeration. Upon isolation, LL12 was propagated using a nonpathogenic *E. coli* K-12 strain, DH5-alpha, by the soft-agar overlay method (3). Phage DNA was purified using a modified Wizard DNA purification kit (Promega) as described previously (4). Phage LL12 DNA was prepared for sequencing as part of a pooled indexed DNA library using the Illumina TruSeq nano low-throughput (LT) kit and sequenced by an Illumina HiSeq 2500 instrument as unpaired 100-base pair reads by sequencing by synthesis (SBS) V2 chemistry; 17,009,166 raw reads generated were quality controlled by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and assembled with Velvet 1.1 (5) into a single contig at 28.3-fold coverage. The contig was closed based on its circular assembly which produced an identical sequence at each end of the contig. Structural annotation was conducted using GLIMMER 3.0 (6) and MetaGeneAnnotator 1.0 (7), with tRNAs predicted by ARAGORN 2.36 (8); gene functions were predicted by InterProScan 5.15-54.0 (9), the NCBI Conserved Domains Database (10), TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM), BLASTp 2.2.8 (11), and HHpred 2.1 (12). Genome annotation was conducted using the Phage Galaxy (13) and WebApollo (14) instances hosted by the Center for Phage Technology (https://cpt.tamu.edu/), and all analyses were conducted using default parameters. Phages were imaged by transmission electron microscopy at the Texas A&M Microscopy and Imaging Center as previously described (15, 16).

The myophage LL12 has a genome of 136,026 bp, with a G+C content of 43.6%. It contains 213 predicted protein-coding genes and 7 tRNAs. Based on BLASTp (E value, $\leq 10^{-5}$), LL12 is most closely related to V5-like myophages, including rV5 (GenBank accession no. NC_011041; 206 shared proteins) and ϕAPCEc02 (KR698074; 204 shared proteins). LL12 is also related to the phage phi92 (NC_023693), with 48 common proteins. An analysis of the Illumina reads by PhageTerm (17) suggests the presence of a nonpermuted terminal redundancy of 459 bp spanning bases 104,966 to 105,424 in

Citation Piya D, Lessor L, Liu M, Gill JJ. 2019. Complete genome sequence of enterotoxigenic *Escherichia coli* myophage LL12. Microbiol Resour Announc 8:e00675-19. https://doi.org/10.1128/MRA.00675-19.

Editor Catherine Putonti, Loyola University Chicago

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.

Received 4 June 2019
Accepted 2 July 2019
Published 25 July 2019

the genome sequence as deposited. This predicted terminal repeat is located in a noncoding region between two convergent transcripts. Fifty LL12-encoded proteins could be assigned putative functions, including the capsid protein (gp59), portal protein (gp63), and large terminase subunit (gp64). The components for tail morphogenesis, including baseplate (gp35) and multiple predicted tail fiber proteins (gp27, gp29, gp32, gp33, gp36, gp41, and gp42), were identified. The genes encoding the large terminase subunit (gp64) and DNA polymerase (gp212) are disrupted by predicted intron sequences. The genes responsible for phage lysis, including the phage endolysin (gp90), i-spanin (gp66), and o-spanin (gp65), were identifiable, but the phage holin could not be positively identified.

Data availability. The annotated phage genome sequence is deposited in NCBI GenBank under accession no. MH491969. Associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR9113738, and SAMN11840293, respectively.

ACKNOWLEDGMENTS

We thank John Deaton of Deerland Enzymes for provision of pathogenic *E. coli* strains.

This work was supported by Deerland Enzymes, Inc., and by funding from the National Science Foundation (awards EF-0949351 and DBI-1565146). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. Additional support came from the Center for Phage Technology (CPT), an Initial University Multidisciplinary Research Initiative supported by Texas A&M University and Texas AgriLife, and from the Texas A&M University Department of Biochemistry and Biophysics.

This announcement was prepared in partial fulfillment of the requirements for BICH464 Phage Genomics.

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. 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.
3. Adams MK. 1959. Bacteriophages. Interscience Publishers, Inc., New York, NY.
4. 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.
5. Zerbino DR. 2010. Using the Velvet *de novo* assembler for short-read sequencing technologies. Curr Protoc Bioinformatics Chapter 11:Unit 11 5. https://doi.org/10.1002/0471250953.bi1105s31.
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. 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.
10. 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.
11. 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.
12. 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.
13. Cock PJ, Gruning BA, Paszkiewicz K, Pritchard L. 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. PeerJ 1:e167. https://doi.org/10.7717/peerj.167.
14. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a Web-based genomic annotation editing platform. Genome Biol 14:R93. https://doi.org/10.1186/gb-2013-14-8-r93.
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. 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.
17. Garneau JR, Depardieu F, Fortier LC, Bikard D, Monot M. 2017. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep 7:8292. https://doi.org/10.1038/s41598-017-07910-5.