Complete Genome Sequence of *Escherichia coli* Myophage Mansfield

Genevieve M. D'Souza,a,b Kathryn Klotz,a Russell Moreland,a Mei Liu,a [ID] Jolene Ramseya

aCenter for Phage Technology, Texas A&M University, College Station, Texas, USA
bDepartment of Animal Science, Texas A&M University, College Station, Texas, USA

ABSTRACT Mansfield is a PB1-like *Escherichia* bacteriophage with a 68,120-bp genome and a predicted 3,673-bp direct terminal repeat. This myophage encodes 105 proteins, for which 32 functions were predicted.

*E*scherichia coli is a commensal Gram-negative bacterium that thrives in the intestinal section of the gastrointestinal tract (1), but some strains are pathogenic (2). *E. coli* O157:H7 is one of the most virulent serotypes, with its strains inducing severe bloody diarrhea and dehydration in infected patients (3). Although humans are a primary host for symbiotic *E. coli* strains, livestock have also been identified as carriers and shedders of strains that are commensal and pathogenic to humans (4). With the rise of antibiotic resistance, bacteriophages are being considered as a precision alternative medicine for eliminating *E. coli* specifically (5). Here, we present the genome sequence of myophage Mansfield, which infects *E. coli*.

Mansfield was isolated from filtered (0.2-μm pore size) stream water in College Station, TX, by propagation on its host, *E. coli* 4s, grown in Luria broth/agar aerobically at 37°C via the soft agar overlay method (6, 7). Mansfield genomic DNA was purified using the Promega Wizard DNA clean-up system described by Summer (8). A library prepared with a TruSeq Nano low-throughput kit was sequenced on an Illumina MiSeq platform using paired-end 250-bp reads with V2 500-cycle chemistry. The 414,121 sequence reads from the index containing the phage genome were quality controlled using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and a complete genome was assembled via SPAdes v3.5.0 (9), with 297.7-fold coverage after trimming with the FASTX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). PCR (forward primer, 5'-CGACACATTCGGTCCACTAA-3'; reverse primer, 5'-TATTGAGCGTTC CCTCGAAAG-3') and Sanger sequencing were used to close the genome. Gene calling was completed with Glimmer v3.0 and MetaGeneAnnotator v1.0 (10, 11). Gene functions were predicted using InterProScan v5.33-72, TMHMM v2.0, and BLAST v2.2.31 at their default settings, cross-referencing hits for BLAST at a 0.001 maximum expected value cutoff versus the NCBI nonredundant, UniProtKB Swiss-Prot, and TrEMBL databases (12–15). Additional evidence came from the HHSuite v3.0 tool HHpred (multiple-sequence alignment [MSA] generation with HHBlits ummiclus30_2018_08 database and modeling with PDB_mmCIF70) (16). TransTermHP v2.09 was used to annotate Rho-independent termination sites (17). The absence of tRNA genes was determined using ARAGORN v2.36 (18). Genome sequence similarities were calculated by progressiveMauve 2.4.0 (19). All tools are hosted in the Center for Phage Technology Galaxy instance, and annotation was performed in Web Apollo (https://cpt.tamu.edu/galaxy-pub) (20, 21). To determine Mansfield's morphology, samples were negatively stained with 2% (wt/vol) uranyl acetate and viewed by transmission electron microscopy at the Texas A&M Microscopy and Imaging Center (22).

The Mansfield genome is 68,120 bp, with a G+C content of 46.14%. The 105

Address correspondence to Jolene Ramsey, jolenerr@tamu.edu.

protein-coding genes are at a 93% coding density, and 32 have a predicted function. Mansfield's genome was opened at the boundary of 3,673-bp direct terminal repeats predicted by PhageTerm (23).

The two phages most closely related to Mansfield are PB1-like phages of the pbunaviruses, *Escherichia* phage ECML-117 (GenBank accession number JX128258) and *Escherichia* phage FEC19 (GenBank accession number MH816966), having 90.89% nucleotide similarity and 90 proteins in common with phage ECML-117, and 90.52% nucleotide similarity with 89 proteins in common for phage FEC19 (24, 25).

Data availability. The genome sequence and associated data for phage Mansfield were deposited under GenBank accession number MK903282, BioProject accession number PRJNA222858, SRA accession number SRR8893603, and BioSample accession number SAMN11414488.

REFERENCES

1. Gomes TAT, Elias WP, Scaletsky ICA, Guth BEC, Rodrigues JF, Piazza RMF, Ferreira LCS, Martinez MB. 2016. Diarrheagenic *Escherichia coli*. Braz J Microbiol 47(Suppl 1):3–30. https://doi.org/10.1016/j.bjm.2016.10.015.
2. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. 2013. Recent advances in understanding enteric pathogenic *Escherichia coli*. Clin Microbiol Rev 26:822–880. https://doi.org/10.1128/CMR.00022-13.
3. Locking ME, Pollock KGJ, Allison LJ, Rae L, Hanson MF, Cowden JM. 2011. *Escherichia coli* O157 infection and secondary spread, Scotland, 1999–2008. Emerg Infect Dis 17:524–527. https://doi.org/10.3201/eid1703.100167.
4. Avery SM, Moore A, Hutchison ML. 2004. Fate of *Escherichia coli* originating from livestock faeces deposited directly onto pasture. Lett Appl Microbiol 38:355–359. https://doi.org/10.1111/j.1472-765X.2004.01501.x.
5. Callaway TR, Edrington TS, Brabban AD, Anderson RC, Rossman ML, Engler MJ, Carr MA, Genovese KJ, Keen JE, Looper ML, Kutter EM, Nisbet DJ. 2008. Bacteriophage isolated from feedlot cattle can reduce *Escherichia coli* O157:H7 populations in ruminant gastrointestinal tracts. Foodborne Pathog Dis 5:183–191. https://doi.org/10.1089/fpd.2007.0057.
6. Golomidova A, Kulikov E, Isaeva A, Manykin A, Letarov A. 2007. The diversity of coliphages and coliforms in horse feces reveals a complex pattern of ecological interactions. Appl Environ Microbiol 73:5975–5981. https://doi.org/10.1128/AEM.01145-07.
7. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
8. 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.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
10. 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.
11. 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.
12. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
13. Krogh A, Larsson B, Heijne von G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
14. 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.
15. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699–2699. https://doi.org/10.1093/nar/gky092.
16. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.
17. Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol 8:R22. https://doi.org/10.1186/gb-2007-8-2-r22.
18. 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.
19. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
20. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
21. 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.
22. 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.

23. Garneau JR, Depardieu F, Fortier L-C, 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.

24. Ceyssens P-J, Miroshnikov K, Mattheus W, Krylov V, Robben J, Noben J-P, Vanderschraeghe S, Sykilinda N, Kropinski AM, Volckaert G, Mesyanzhinov V, Lavigne R. 2009. Comparative analysis of the widespread and conserved PB1-like viruses infecting Pseudomonas aeruginosa. Environ Microbiol 11: 2874–2883. https://doi.org/10.1111/j.1462-2920.2009.02030.x.

25. Watkins SC, Sible E, Putonti C. 2018. Pseudomonas PB1-like phages: whole genomes from metagenomes offer insight into an abundant group of bacteriophages. Viruses 10:331. https://doi.org/10.3390/v10060331.