



## Complete Genome Sequence of *Escherichia coli* Myophage Mangalitsa

Cameron L. Atkison,<sup>a</sup> Justin Boeckman,<sup>a,b</sup> Heather Newkirk,<sup>a</sup> Mei Liu,<sup>a</sup> DJason J. Gill,<sup>a,b</sup> Jesse Cahill,<sup>a\*</sup> Jolene Ramsey<sup>a</sup>

<sup>a</sup>Center for Phage Technology, Texas A&M University, College Station, Texas, USA <sup>b</sup>Department of Animal Science, Texas A&M University, College Station, Texas, USA

**ABSTRACT** Enteropathogenic *Escherichia coli* is a prevalent Gram-negative bacterium that can lead to fatal complications from infection in humans. Here, we present the isolation and complete annotation of the 52,329-bp genome of enteropathogenic *E. coli* ATCC 23545 myophage Mangalitsa. Predicted terminal repeats and temperature sensitivity for plaque formation place Mangalitsa with similar unclassified myoviruses.

Enteropathogenic *Escherichia coli* (EPEC) strains are Gram-negative human pathogens that are prevalent in both communal and clinical settings (1). They have been identified as one of the leading causes of persistent diarrhea, which is the second largest contributor to childhood mortality, accounting for 1.3 million deaths per year (2, 3). Here, we present the complete genome sequence of enteropathogenic *E. coli* myophage Mangalitsa.

Bacteriophage Mangalitsa was isolated from a chloroform-sterilized and enriched swine fecal sample collected in College Station, TX, based on its ability to grow on the enteropathogenic E. coli strain ATCC 23545. While the host was typically grown aerobically at 37°C in Luria broth (BD) and standard soft agar overlay methods were used (4), Mangalitsa only produced plaques at 30°C or 22°C. Phage genomic DNA was isolated using the shotgun library prep modification of the Promega Wizard DNA clean-up system (5). A genomic library prepared with the TruSeq Nano low-throughput kit was sequenced by an Illumina MiSeq platform with paired-end 250-bp reads. A total of 480,501 reads were in the phage index. Quality control was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/). Sequence reads were then trimmed using the FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx\_toolkit/). The genome was assembled with 1,333-fold coverage using SPAdes v3.5.0 and closed by PCR (forward primer, 5'-AGTGCACGGTAT TCTTCGTTAG-3'; reverse primer, 5'-CTAACGCATCGAATCTCTTCTCA-3') and Sanger sequencing (6). Structural annotations were made with GLIMMER v3.0 and MetaGene-Annotator v1.0, and ARAGORN v2.36 did not reveal any tRNAs (7-9). Protein-coding gene function was predicted with InterProScan v5.33-72 and BLAST v2.2.31 (10, 11). The BLAST analysis queried the NCBI nonredundant and UniProtKB Swiss-Prot and TrEMBL databases with a 0.001 maximum expectation value cutoff (12). Rho-independent termination sites were analyzed using TransTermHP v2.09 (13). Whole-genome similarity was calculated by the progressiveMauve v2.4.0 algorithm (14). The annotation tools used are in the Galaxy and Web Apollo instances hosted by the Center for Phage Technology (https://cpt.tamu.edu/galaxy-pub). The morphology of the phage sample was determined by negative staining with 2% uranyl acetate and visualized with transmission electron microscopy at the Texas A&M University Microscopy and Imaging Center (15).

Citation Atkison CL, Boeckman J, Newkirk H, Liu M, Gill JJ, Cahill J, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* myophage Mangalitsa. Microbiol Resour Announc 8:e01045-19. https://doi.org/10.1128/ MRA.01045-19.

Editor John J. Dennehy, Queens College Copyright © 2019 Atkison et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Jesse Cahill, jesse.cahill@tamu.edu, or Jolene Ramsey, jolenerr@tamu.edu.

\* Present address: Jesse Cahill, Sandia National Laboratories, Albuquerque, New Mexico, USA.

Received 26 August 2019 Accepted 30 August 2019 Published 19 September 2019 Mangalitsa is a myophage with a 52,329-bp genome consisting of 44.01% GC content, 1 terminator, 0 tRNAs, and 82 predicted protein-coding genes. Of the protein-coding genes, 31 have a predicted function, while 51 are hypothetical proteins, totaling an overall 89.66% coding density. Additionally, Mangalitsa has 3,234-bp terminal repeats detected by PhageTerm (16). The terminal repeat pattern is consistent with its most closely related phages, which include enterobacterial phage phiEcoM-GJ1 (79.61% nucleotide similarity; GenBank accession no. EF460875), *Pectobacterium* phage PP101 (47.01% nucleotide similarity; KY087898), *Pectobacterium* phage PM1 (46.24% nucleotide similarity; KF534715), and *Erwinia* phage vB\_EamM-Y2 (37.28% nucleotide similarity; HQ728264) (17–19). As such, the Mangalitsa genome was reopened with the RNA polymerase gene as the first feature following the terminal repeat to follow the convention for this group of phages. Interestingly, temperature sensitivity was also reported for the related *Escherichia* myophage ST32 (MF044458) (20).

**Data availability.** The genome sequence and associated data for phage Mangalitsa were deposited under GenBank accession no. MN045229, BioProject no. PRJNA222858, SRA no. SRR8869233, and BioSample no. SAMN11360419.

## **ACKNOWLEDGMENTS**

This work was supported by funding from the National Science Foundation (awards EF-0949351 and DBI-1565146) and the Bill and Melinda Gates Foundation (project OPP1139800). 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 Department of Biochemistry and Biophysics of Texas A&M University.

We are grateful for the advice and support of the CPT staff.

This announcement was prepared in partial fulfillment of the requirements for BICH464 Bacteriophage Genomics, an undergraduate course at Texas A&M University.

## REFERENCES

- Ochoa TJ, Contreras CA. 2011. Enteropathogenic Escherichia coli infection in children. Curr Opin Infect Dis 24:478–483. https://doi.org/10 .1097/QCO.0b013e32834a8b8b.
- Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, Eisele T, Liu L, Mathers C, Child Health Epidemiology Reference Group of WHO and UNICEF. 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 375:1969–1987. https://doi.org/10.1016/S0140 -6736(10)60549-1.
- Abba K, Sinfield R, Hart CA, Garner P. 2009. Pathogens associated with persistent diarrhoea in children in low and middle income countries: systematic review. BMC Infect Dis 9:88. https://doi.org/10.1186/1471 -2334-9-88.
- 4. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
- 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.
- 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.
- 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.
- 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.
- 9. 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.

- 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.
- 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.
- UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/ gky1049.
- 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.
- 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.
- 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.
- 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.
- Jamalludeen N, Kropinski AM, Johnson RP, Lingohr E, Harel J, Gyles CL. 2008. Complete genomic sequence of bacteriophage φEcoM-GJ1, a novel phage that has myovirus morphology and a podovirus-like RNA polymerase. Appl Environ Microbiol 74:516–525. https://doi.org/10 .1128/AEM.00990-07.

- Lim J-A, Shin H, Lee DH, Han S-W, Lee J-H, Ryu S, Heu S. 2014. Complete genome sequence of the Pectobacterium carotovorum subsp. carotovorum virulent bacteriophage PM1. Arch Virol 159:2185–2187. https:// doi.org/10.1007/s00705-014-2005-7.
- Born Y, Fieseler L, Marazzi J, Lurz R, Duffy B, Loessner MJ. 2011. Novel virulent and broad-host-range Erwinia amylovora bacteriophages reveal

a high degree of mosaicism and a relationship to Enterobacteriaceae phages. Appl Environ Microbiol 77:5945–5954. https://doi.org/10.1128/ AEM.03022-10.

 Liu H, Geagea H, Rousseau GM, Labrie SJ, Tremblay DM, Liu X, Moineau S. 2018. Characterization of the Escherichia coli virulent myophage ST32. Viruses 10:616. https://doi.org/10.3390/v10110616.