AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Escherichia coli* Podophage Peacock

Erin Ruhlman,[a] Ryan Bockoven,[a]* Russell Moreland,[a] Mei Liu,[a] Jolene Ramsey[a]

[a]Center for Phage Technology, Texas A&M University, College Station, Texas, USA

ABSTRACT *Escherichia coli* is typically a commensal bacterium of the mammalian intestinal tract. Here, the isolation and annotation of the 39,233-bp T7-like *E. coli* podophage Peacock genome are described.

Escherichia coli is a Gram-negative, rod-shaped bacterium found in the mammalian intestinal tract. It has been widely studied due to its high growth rate and genetic plasticity (1). Although *E. coli* is typically a commensal organism, pathogenic strains can arise, for instance, through the acquisition of virulence factors through horizontal gene transfer (2). Antibiotics remain the standard of treatment even though many strains have become resistant (3). Investigating *E. coli* bacteriophages might provide insight into potential alternative treatments.

Phage Peacock was isolated from filtered (0.2-μm pore size) wastewater treatment plant influent samples collected in College Station, TX, using *E. coli* 4s as a host (4). The host was cultured aerobically on lysogeny broth/agar at 37°C with the soft agar overlay method, and genomic DNA was extracted from this phage, as previously described, using the shotgun library preparation modification to the Promega Wizard DNA clean-up system (5, 6). A Peacock DNA library was prepared with the TruSeq Nano low-throughput kit and sequenced on an Illumina MiSeq platform with paired-end 250-bp reads using V2 500-cycle chemistry. The 570,679 total sequence reads from the index containing the phage genome were evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using the FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). The Peacock genome was assembled in a single raw contig with SPAdes v3.5.0 and had 513.9-fold coverage (7). PCR across the contig ends (forward primer 5′-TGTATGGGTATCATCGGG ACA-3′ and reverse primer 5′-CCTCCTTGGACTTAGGGTCATA-3′) and Sanger sequencing were used to manually verify that the correct and complete phage sequence was present in the assembly. All annotation tools described below are in Galaxy and Web Apollo, hosted by the Center for Phage Technology (https://cpt.tamu.edu/galaxy-pub) (8, 9). Gene calling for the genome was completed using Glimmer v3.0 (10) and MetaGeneAnnotator v1.0 (11). No tRNAs were detected using ARAGORN v2.36 (12). Gene functions were predicted using InterProScan v5.33-72 (13), BLAST v2.2.31 at a 0.001 maximum expectation value (14), and TMHMM v2.0 (15) at default settings. The NCBI nonredundant and UniProtKB Swiss-Prot/TrEMBL databases were used in BLAST functional analysis (16). LipoP v1.0 was used to assess lipobox presence in spanins (17). Rho-independent termination sites were annotated from TransTermHP v2.09 (18). Genome-wide DNA sequence similarity was calculated using progressiveMauve v2.4.0 (19). Unless otherwise stated, all tools were executed using default parameters. To determine phage morphology, samples were negatively stained with 2% (wt/vol) uranyl acetate and viewed using transmission electron microscopy at the Texas A&M Microscopy and Imaging Center (20).

The 39,233-bp genome of the podophage Peacock has a G+C composition of 50.1%

Citation Ruhlman E, Bockoven R, Moreland R, Liu M, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* podophage Peacock. Microbiol Resour Announc 8:e01056-19. https://doi.org/10.1128/MRA.01056-19.

Editor Catherine Putonti, Loyola University Chicago

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

* Present address: Ryan Bockoven, Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas, USA.

Received 26 August 2019
Accepted 30 August 2019
Published 26 September 2019

and a 92.4% coding density. Of the 55 protein-coding genes identified, 30 functions were predicted, and 27 of those had direct BLASTp hits to phage T7 (GenBank accession number NC_001604). Using PhageTerm, Peacock was predicted to have 179-bp direct terminal repeats (21). Peacock has 85% nucleotide sequence identity with *Escherichia* phage Vec13 (GenBank accession number MH400309), a T7-like phage, and shares 47 proteins with this phage. An overcome classical restriction (ocr) protein (NCBI accession number QEG09667), known to inhibit type I DNA restriction enzymes by mimicking a B-form DNA structure, was also identified (22).

Data availability. The genome sequence and associated data for phage Peacock were deposited under GenBank accession number MK903279, BioProject accession number PRJNA222858, SRA accession number SRR8892143, and BioSample accession number SAMN11408679.

ACKNOWLEDGMENTS

This work was supported by funding from the National Science Foundation (award DBI-1565146). 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 thank A. Letarov for the kind gift of the *Escherichia coli* strain 4s. 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

1. Jang J, Hur H-G, Sadowsky MJ, Byappanahalli MN, Yan T, Ishii S. 2017. Environmental Escherichia coli: ecology and public health implications–a review. J Appl Microbiol 123:570–581. https://doi.org/10.1111/jam.13468.
2. 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.
3. Yang S-C, Lin C-H, Aljuffali IA, Fang J-Y. 2017. Current pathogenic Escherichia coli foodborne outbreak cases and therapy development. Arch Microbiol 199:811–825. https://doi.org/10.1007/s00203-017-1393-y.
4. 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.
5. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
6. 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.
7. 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.
8. 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.
9. 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.
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. 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.
13. 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.
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. 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.
16. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
17. Juncker AS, Willenbrock H, Heijne Von G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662. https://doi.org/10.1110/ps.0303703.
18. 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.
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. 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.
21. 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.
22. Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, Edwardson JM, Dryden D. 2002. Structure of ocr from bacteriophage T7, a protein that mimics B-form DNA. Mol Cell 9:187–194. https://doi.org/10.1016/S1097-2765(02)00435-5.