AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Escherichia coli* Podophage Penshu1

Douglas Pechacek,[a] Myung Hwangbo,[a,b] Russell Moreland,[a] Mei Liu,[a] Jolene Ramsey[a]

[a]Center for Phage Technology, Texas A&M University, College Station, Texas, USA
[b]Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas, USA

ABSTRACT *Escherichia coli* 4s is a Gram-negative bacterium found in the equine intestinal ecosystem alongside diverse other coliform bacteria and bacteriophages. This announcement describes the complete genome of the T7-like *E. coli* 4s podophage Penshu1. From its 39,263-bp genome, 54 protein-encoding genes and a 179-bp terminal repeat were predicted.

Escherichia coli is a Gram-negative bacterium found living among the intestinal microbiome of all mammals. Due to multiple protective abilities, including extreme acid resistance, *E. coli* colonizes the intestine as a commensal (1). *E. coli* 4s was isolated from horse feces and lives among a large diversity of coliform bacteria in the equine gut ecosystem (2). Enteric bacteriophages significantly influence the bacterial composition and exert pathogen suppression (2–4). Here, we describe a newly isolated *E. coli* podophage called Penshu1.

Penshu1 was isolated from a filtered (filter size, 0.2 μm) wastewater treatment sample from Bryan, TX, using *E. coli* 4s as the host (2). The phage was propagated using the soft-agar overlay method in Luria broth (BD) under aerobic conditions at 37°C (5). Following isolation, Penshu1 podophage morphology was observed using 2% (wt/vol) uranyl acetate negative staining and transmission electron microscopy performed at the Texas A&M University Microscopy and Imaging Center (6). The phage genomic DNA was extracted as described previously (shotgun library preparation modification of the Promega Wizard DNA clean-up system), and libraries were prepared using an Illumina TruSeq Nano low-throughput kit (7). The DNA was sequenced with an Illumina MiSeq platform as paired-end 250-bp reads using v2 500-cycle chemistry. The resulting 565,076 sequence reads from the index containing the phage genome were quality controlled by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After trimming using FastX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/), they were assembled into a single contig at 376.4-fold coverage using SPAdes v3.5.0 (8). Contig completion was confirmed by PCR (forward primer, 5′-TGAAGTCTCATGCACTC TTTCC-3′; reverse primer, 5′-CCCTCGTCTATCTTGTGGAATC-3′) and by Sanger sequencing of the resulting product. All of the tools listed here for assembly and annotation were used at default parameters and are available in the Center for Phage Technology Galaxy instance with integrated Web Apollo (https://cpt.tamu.edu/galaxy-pub/) (9, 10). Protein-coding genes were predicted using GLIMMER v3.0 and MetaGeneAnnotator v1.0 (11, 12). No tRNA genes were found after analysis with ARAGORN v.2.36 (13). Protein-coding gene functions were predicted using BLAST v2.2.31 with a 0.001 maximum expectation value, LipoP v1.0, and TMHMM v2.0, and conserved domains were found using InterProScan v5.33-72 (14–17). Rho-independent termination sites were identified using TransTermHP v2.09 (18). Genome-wide DNA sequence similarity comparisons were carried out with progressiveMauve v2.4.0 (19).

Penshu1 has a 39,263-bp genome, with a 93.4% coding density and 50.6% G+C

Citation Pechacek D, Hwangbo M, Moreland R, Liu M, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* podophage Penshu1. Microbiol Resour Announc 8:e01055-19. https://doi.org/10.1128/MRA.01055-19.

Editor Simon Roux, DOE Joint Genome Institute

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

Received 26 August 2019
Accepted 3 September 2019
Published 19 September 2019

content. Analysis predicted 54 protein-coding genes, with 30 being assigned a putative function. The Penshu1 genome was reopened at T7-like direct terminal repeats of 179 bp predicted by PhageTerm (20). Penshu1 has its highest identity with several unclassified T7-like phages, including 43 similar proteins and 80.3% nucleotide sequence identity with *Escherichia* phage ST31 (GenBank accession number KY962008) and 80.1% nucleotide sequence identity with *Escherichia* phage YZ1 (GenBank accession number MG845865). As for phage T7, Penshu1 has a slippery sequence in the major capsid protein (NCBI accession number QEG09806) that can lead to translation by frameshift of the minor capsid protein (NCBI accession number QEG09807).

Data availability. The genome sequence and associated data for phage Penshu1 were deposited under GenBank accession number MK903281, BioProject accession number PRJNA222858, SRA accession number SRR8893626, and BioSample accession number SAMN11414580.

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 Texas A&M University Department of Biochemistry and Biophysics.

We thank A. Letarov for the kind gift of *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. Foster JW. 2004. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nat Rev Microbiol 2:898–907. https://doi.org/10.1038/nrmicro1021.
2. 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.
3. Knirel YA, Prokhorov NS, Shashkov AS, Ovchinnikova OG, Zdorovenko EL, Liu B, Kostryukova ES, Larin AK, Golomidova AK, Letarov AV. 2015. Variations in O-antigen biosynthesis and O-acetylation associated with altered phage sensitivity in *Escherichia coli* 4s. J Bacteriol 197:905–912. https://doi.org/10.1128/JB.02398-14.
4. Daly K, Stewart CS, Flint HJ, Shirazi-Beechey SP. 2001. Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. FEMS Microbiol Ecol 38:141–151. https://doi.org/10.1111/j.1574-6941.2001.tb00892.x.
5. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
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. Juncker AS, Willenbrock H, Von Heijne 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.
16. 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.
17. 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.
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. 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.