AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Escherichia coli* Siphophage Schulenberg

Madeline Rivera,[a,b] Heather Newkirk,[a] Russell Moreland,[a] Mei Liu,[a] Jolene Ramsey[a]

[a]Center for Phage Technology, Texas A&M University, College Station, Texas, USA
[b]Department of Animal Science, Texas A&M University, College Station, Texas, USA

ABSTRACT *Escherichia coli* bacteria and their infecting bacteriophage exist within the gut. Here, we present the complete genome of Schulenberg, an *E. coli* siphophage similar to phages of the subfamily *Guernseyvirinae*. Schulenberg encodes 85 proteins, 33 of which have predicted functions.

Escherichia coli is a Gram-negative facultative anaerobe commonly found in the lower gastrointestinal tract of mammals (1). Diversity within gut-dwelling bacterial populations aids in digestion and maintenance of healthy immune function. Some groups have proposed that phage therapy manipulating gut microbiota may prove useful in treating chronic diseases in humans (2). Here, we report the *E. coli*-infecting siphophage Schulenberg.

Bacteriophage Schulenberg was isolated on host *E. coli* 4s from a filtered (0.2-μm pore size) private septic system sample collected in Franklin, TX (3). Both phage and host were grown aerobically at 37°C in Luria broth (BD), and standard soft agar overlay methods were used (4). Phage morphology was determined using transmission electron microscopy at the Texas A&M Microscopy and Imaging Center after staining with 2% (wt/vol) uranyl acetate (5). Phage genomic DNA libraries were prepared after isolation with the shotgun library preparation modifications to the Promega Wizard DNA clean-up system using Illumina TruSeq Nano low-throughput kits and were sequenced with an Illumina MiSeq instrument with 250-bp paired-end reads using V2 500-cycle chemistry (6). The 4,413 raw sequence reads were quality controlled with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and then trimmed using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Genomes were assembled with SPAdes v3.5.0 at default settings into a single contig of circular assembly at 7.1-fold coverage (7). The product was confirmed and closed using PCR (forward primer, 5′-GAGAAGTTACGAGAGTACAGGAGTAATA-3′; reverse primer, 5′-GCCAACACCTTCTCCA TCT-3′) and Sanger sequencing of the product. Glimmer v3.0 and MetaGeneAnnotator v1.0 and ARAGORN v2.36 were used to call protein-coding and tRNA genes, respectively (8–10). Rho-independent termination sites were annotated from TransTermHP v2.09 (11). Gene functions were predicted using domain scans with InterProScan v5.33-72, TMHMM v2.0, and BLAST v2.2.31 at default parameters (with a 0.001 maximum expectation value) for the NCBI nonredundant and UniProtKB Swiss-Prot/TrEMBL databases (12–15). Structural similarity predictions were carried out with the HHSuite v3.0 HHpred tool (multiple-sequence alignment [MSA] generation with the HHblits ummiclus30_2018_08 database and modeling with PDB_mmCIF70) (16). Sequence similarity to other phages was calculated using progressiveMauve v2.4.0 (17). All tools used for annotations are available on the Center for Phage Technology Galaxy and Web Apollo instances (http://cpt.tamu.edu/galaxy-pub/) (18, 19).

The 44,748-bp double-stranded DNA genome of siphophage Schulenberg was

Citation Rivera M, Newkirk H, Moreland R, Liu M, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* siphophage Schulenberg. Microbiol Resour Announc 8:e01053-19. https://doi.org/10.1128/MRA.01053-19.

Editor John J. Dennehy, Queens College

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

Received 26 August 2019
Accepted 4 September 2019
Published 26 September 2019

predicted to be packaged by a headful mechanism using PhageTerm (20). Schulenberg has 85 predicted protein-coding genes, 33 of which have a predicted function, no identifiable tRNA genes, a coding density of 94.8%, and a G+C content of 49.9%. The most closely related phage to Schulenberg is *Escherichia* phage VB_EcoS-Golestan (GenBank accession number MG099933) within the subfamily *Guernseyvirinae*, with 67.63% nucleotide similarity and 58 similar proteins (21). Unlike other phages in the *Guernseyvirinae* subfamily, Schulenberg contains no detectable inteins, but it carries two freestanding HNH endonucleases (NCBI accession numbers QEG06793 and QEG06824). A superinfection immunity protein (NCBI accession number QEG06814) with similarity to the *Escherichia* phage T4 Imm protein (NCBI accession number NP_049660) was also found. The Schulenberg holin/antiholin pair (NCBI accession numbers QEG06859 and QEG06860) and endolysin (NCBI accession number QEG06861) are genetically separate from the partially overlapping i-spanin/o-spanin (NCBI accession numbers QEG06797 and QEG06798).

Data availability. The genome sequence and associated data for phage Schulenberg have been deposited under GenBank accession number MK931438, BioProject accession number PRJNA222858, SRA accession number SRR8892142, and BioSample accession number SAMN11408657.

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, 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. Blount ZD. 2015. The natural history of model organisms: the unexhausted potential of E. coli. Elife 4:e05826. https://doi.org/10.7554/eLife.05826.
2. Paule A, Frezza D, Edeas M. 2018. Microbiota and phage therapy: future challenges in medicine. Med Sci (Basel) 6:86. https://doi.org/10.3390/medsci6040086.
3. 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.
4. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
5. 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.
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. 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.
9. 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.
10. 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.
11. 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.
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. 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. 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.
18. 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.
19. 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.
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.
21. Anany H, Switt AIM, De Lappe N, Ackermann H-W, Reynolds DM, Kropinski AM, Wiedmann M, Griffiths MW, Tremblay D, Moineau S, Nash JHE, Turner D. 2015. A proposed new bacteriophage subfamily: "Jerseyvirinae." Arch Virol 160:1021–1033. https://doi.org/10.1007/s00705-015-2344-z.