AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Escherichia coli* Siphophage Shashou

Tyler Higbee,[a] Shih-Hung Yang,[a] Russell Moreland,[a] Mei Liu,[a] Jolene Ramsey[a]

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

ABSTRACT Here, we announce the genome of the *Escherichia coli* 4s siphophage Shashou, which presents similarity to members of the *Guernseyvirinae* subfamily. Shashou is predicted to use a headful packaging mechanism for its 44,155-bp genome and to encode 77 proteins.

Escherichia coli, a Gram-negative bacterium with fast growth under diverse conditions, is one of the most studied organisms on the planet (1). Primarily, *E. coli* inhabits the gut biome of mammals and, less frequently, soil, water, and food. Phages that infect *E. coli* may be useful in phage therapy against certain antibiotic-resistant pathogenic strains (2). The well-studied host provides a robust experimental system for studying the phage life cycle. This announcement reports the genome of the new *E. coli* 4s siphophage Shashou.

Phage Shashou was isolated from filtered (0.2-μm filter) wastewater treatment plant influent collected in College Station, Texas, by plating on *Escherichia coli* 4s (3). Both phage and host were grown aerobically at 37°C in Luria broth (BD), and standard soft-agar overlay methods were used (4). Genomic DNA was purified using the Promega Wizard DNA clean-up system according to the modification in the shotgun library preparation protocol given by Summer and prepared as Illumina TruSeq Nano low-throughput libraries (5). Sequencing was done with paired-end 250-bp reads using V2 500-cycle chemistry on an Illumina MiSeq system. The 844,502 total sequence reads from the index containing the phage genome were quality controlled using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were assembled with SPAdes v3.5.0 using default parameters, yielding a single contig of 44,155 bp for Shashou and 402.6-fold contig coverage after trimming using FastX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) (6). Sanger sequencing of a PCR product amplified across the 5′ and 3′ contig ends (forward, 5′-GGACTCTATATGTCAAGCG GATG-3′; reverse, 5′-TGGCAGGAAATTACAGCGTAG-3′) was used to close the genome. Rho-independent termination sites were annotated from TransTermHP v2.09 (7). tRNA detection was done with ARAGORN v2.36, and gene calling was performed using GLIMMER v3.0 and MetaGeneAnnotator v1.0 (8–10). TMHMM v2.0, InterProScan v5.33-72, and BLAST v2.2.31 searches against the NCBI nonredundant (nr) and UniProtKB TrEMBL and Swiss-Prot databases with a 0.001 minimum expectation cutoff were used to predict gene function (11–14). HHPred, using the HHsuite v3.0 release at default settings with HHblits with ummiclust30_2018_08 for multiple-sequence alignment (MSA) generation and PDB_mmCIF70 for modeling, was used to predict structural similarity (15). DNA sequence similarity was calculated using progressiveMauve v2.4.0 (16). The Galaxy and Web Apollo annotation tools are hosted at the Center for Phage Technology (https://cpt.tamu.edu/galaxy-pub) (17, 18). Phage samples were stained with 2% (wt/vol) uranyl acetate and viewed using transmission electron microscopy at the Texas A&M Microscopy and Imaging Center for morphology observations (19).

Shashou is a siphophage with a 44,155-bp genome with 77 predicted protein-

Citation Higbee T, Yang S-H, Moreland R, Liu M, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* siphophage Shashou. Microbiol Resour Announc 8:e01016-19. https://doi.org/10.1128/MRA.01016-19.

Editor John J. Dennehy, Queens College

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

Received 20 August 2019
Accepted 13 September 2019
Published 3 October 2019

coding genes. It has a 50.1% G+C content and 94.4% coding density. PhageTerm predicts that Shashou uses a pac site for packaging (20).

Shashou shares 65.15% nucleotide identity and 51 proteins with *Escherichia* phage G AB-2017 (GenBank accession number KY295895), a *Guernseyvirinae* subfamily member (21). A hypothetical protein was predicted between the tail assembly chaperone (NCBI accession number QEA09424) and the tape measure protein (NCBI accession number QEA09426), with no canonical slippery sequence to induce translational frameshifting. The helicase (NCBI accession number QEA09433) in Shashou contains an intein.

Data availability. The genome sequence and associated data for phage Shashou were deposited under GenBank accession number MK931440, BioProject accession number PRJNA222858, SRA accession number SRR8893605, and BioSample accession number SAMN11414490.

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 4:96. https://doi.org/10.7554/eLife.05826.
2. Brüssow H. 2005. Phage therapy: the Escherichia coli experience. Microbiology 151:2133–2140. https://doi.org/10.1099/mic.0.27849-0.
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. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
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. 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.
14. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699–2699. https://doi.org/10.1093/nar/gky092.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
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.