AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Escherichia coli* Siphophage Snoke

James E. Corban,[a] Jacob Gramer,[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 a Gram-negative bacterium often found in animal intestinal tracts. Here, we present the genome of the *Guernseyvirinae*-like *E. coli* 4s siphophage Snoke. The 44.4-kb genome contains 81 protein-coding genes, for which 33 functions were predicted. The capsid morphogenesis gene in Snoke contains a large intein.

Escherichia coli, a heavily studied Gram-negative bacterium, is most commonly found in the intestinal tract of humans and other animals (1). Selective pressures generated from bacteriophage-bacterium interactions in the gut shape the development of both organisms (2). Studying bacteriophages that grow on *E. coli* 4s (3), a horse fecal isolate, may illuminate the mechanisms of phage-bacterium coevolution within unique natural microbiomes (4). Here, we present the complete genome sequence of the *E. coli* 4s siphophage Snoke.

Bacteriophage Snoke was isolated from filtered (0.2-μm-pore-size) activated sludge sourced from a wastewater treatment facility in Austin, TX. The phage was propagated on *E. coli* 4s aerobically at 37°C in Luria broth (BD) using the soft-agar overlay methods described in reference 5. Genomic DNA was purified from the phage as previously described with the Promega Wizard DNA clean-up system (6), prepared as Illumina TruSeq Nano low-throughput libraries, and sequenced in paired-end 250-bp reads using v2 500-cycle chemistry on an Illumina MiSeq platform. The 565,076 sequence reads from the index containing the phage genome were quality controlled with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the phage genome was assembled into a single raw contig via SPAdes v.3.5.0 with 1,291.6-fold coverage after trimming using the FASTX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) (7). PCR amplification across the raw contig ends (forward primer 5′-GCAACACGGCACAGAAAC-3′ and reverse primer 5′-CTGCGACGGAGAAATCAACT-3′) and verification by Sanger sequencing of the DNA product ensured that the contig sequence was complete. Accompanied with manual corrections, protein-coding gene predictions were performed using GLIMMER v3.0 and MetaGeneAnnotator v1.0; predicted gene functions were assigned using InterProScan v5.33-72, the HHSuite v3.0 HHpred tool (multiple sequence alignment [MSA] generation with the HHblits ummiclus30_2018_08 database and modeling with PDB_mmCIF70), BLAST v2.2.31 with a 0.001 maximum expectation value, and TMHMM v2.0 at the default settings (8–13). All BLAST queries were run against the NCBI nonredundant and UniProtKB Swiss-Prot and TrEMBL databases (14). ARAGORN v2.36 was run to detect tRNA coding sequences (15). Rho-independent termination sites were annotated using TransTermHP v2.09 (16). DNA sequence similarity was determined using the progressiveMauve v2.4.0 alignment algorithm (17). The genome annotation tools (with the exception of HHpred) were accessed via the Center for Phage Technology Galaxy and Web Apollo interfaces (https://cpt.tamu.edu/galaxy-pub) and run with default parameters, unless stated above (18, 19). The morphology of bacteriophage Snoke was determined using samples

Citation Corban JE, Gramer J, Moreland R, Liu M, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* siphophage Snoke. Microbiol Resour Announc 8:e01051-19. https://doi.org/10.1128/MRA.01051-19.

Editor Catherine Putonti, Loyola University Chicago

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

Received 26 August 2019
Accepted 30 August 2019
Published 3 October 2019

negatively stained with 2% (wt/vol) uranyl acetate and viewed by transmission electron microscopy at the Texas A&M Microscopy and Imaging Center (20).

The 44,454-bp genome of Snoke has a G+C content of 50.2% and 95.0% coding density. Genome analysis indicates 81 protein-coding genes, of which 33 received a functional annotation, but no tRNA genes. The Snoke genome is predicted by PhageTerm to be packaged via a pac-type (headful) DNA packaging mechanism (21). Snoke shares 69.0% nucleotide sequence identity and 58 proteins with *Escherichia* phage VB_EcoS-Golestan (GenBank accession no. MG099933), a member of the *Guernseyvirinae* subfamily (22).

A large self-splicing intein resides within the capsid morphogenesis protein (NCBI accession no. QEG06950) (23). The capsid morphogenesis protein in *Salmonella* phage LSPA1 (GenBank accession no. KM272358) contains a nearly identical intein, and K1ind2 (GenBank accession no. GU196280) lacks the intein.

Data availability. The genome sequence and associated data for phage Snoke were deposited under GenBank accession no. MK931441, BioProject accession no. PRJNA222858, SRA accession no. SRR8892143, and BioSample accession no. 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. Scanlan PD. 2017. Bacteria-bacteriophage coevolution in the human gut: implications for microbial diversity and functionality. Trends Microbiol 25:614–623. https://doi.org/10.1016/j.tim.2017.02.012.
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. Babenko VV, Golomidova AK, Ivanov PA, Letarova MA, Kulikov EE, Manolov AI, Prokhorov NS, Kostrukova ES, Matyushkina DM, Prilipov AG, Maslov S, Belalov IS, Clokie M, Letarov AV. 2019. Phages associated with horses provide new insights into the dominance of lateral gene transfer in virulent bacteriophages evolution in natural systems. bioRxiv. https://doi.org/10.1101/542787.
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. 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. 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.
11. 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.
12. 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.
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. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
15. 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.
16. 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.
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. 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. 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.
23. Shah NH, Muir TW. 2014. Inteins: nature's gift to protein chemists. Chem Sci 5:446–461. https://doi.org/10.1039/C3SC52951G.