AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequences of Mycobacteriophages Candle, Schatzie, Sumter, and Waleliano

Kayla M. Fast,[a] Brianna E. Forrest,[a] Garren R. Granec,[a] John D. Larrimore,[a] Anna E. Morse,[a] Emma D. Ryan,[a] P. Kiersten Schellhammer,[a] Tracy W. Keener,[a] Michael W. Sandel[a]

[a]Department of Biological and Environmental Sciences, University of West Alabama, Livingston, Alabama, USA

ABSTRACT Mycobacteriophages Candle, Schatzie, Sumter, and Waleliano were isolated from soil using the host bacterium *Mycobacterium smegmatis* mc²155. Candle, Schatzie, and Sumter were discovered in Alabama and Waleliano in Maryland. The bacteriophages have been assigned clusters based on nucleotide similarity, as follows: Candle, R; Schatzie, J; Sumter, A1; and Waleliano, B4.

It is estimated that bacteriophage particles are some of the most abundant biological entities on the planet (1). Four bacteriophages (phages) were discovered and the genomes annotated by University of West Alabama (UWA) students and faculty. The phages, designated Candle, Schatzie, Sumter, and Waleliano, were discovered in the United States (Table 1).

The host bacterium *Mycobacterium smegmatis* mc²155 was used in the isolation of these phages. *M. smegmatis* was cultured from a frozen glycerol stock with incubation in 7H9 medium at 37°C with shaking (250 rpm) for 5 days (2). For our purposes of characterizing and finding novel phages, this host is suitable because data are available for over 10,000 phages known to infect *M. smegmatis* (as of June 2019 [3]). Candle and Schatzie were isolated from soil samples at 37°C directly in 7H9 medium (2). Sumter and Waleliano were enriched with the host bacterium in addition to incubation in 7H9 medium. Imaging with a transmission electron microscope showed that all four phages belong to the family *Siphoviridae,* given the noncontractile tails, which ranged from 142.0 to 335.29 nm long, and with isometric heads ranging from 42.0 to 83.3 nm.

Purified high-titer lysates (HTLs) were collected from plaques, and the phage life cycles are given in Table 1. Whole-genomic DNA was extracted from HTLs using the Wizard DNA clean-up system (Promega, Madison, WI [2]). The NEBNext Ultra II FS kit (New England BioLabs, Ipswich, MA) with dual-indexed barcoding was used to build DNA libraries, which were then pooled for sequencing. Sequencing was performed at the Pittsburgh Bacteriophage Institute using an Illumina MiSeq platform. Each genome yielded at least 200,000 single-end 150-base reads; the coverage depths are listed in Table 1. Assembly was performed using Newbler 2.9, with default settings (4). This produced a single contig for each genome, which was used to determine genome ends and was checked for completeness and accuracy using Consed 2.0 (5). Two genomes exhibited defined ends with a 4-bp overhang (ATCC) in Schatzie and 10-bp overhang (CGGATGGTAA) in Sumter. Candle and Waleliano did not show a buildup of reads or coverage variation and were designated circularly permuted, as previously described (6). The beginning of each genome was chosen based on similar phages. The genome lengths and G+C contents are given in Table 1.

Genome annotation was performed in DNA Master 5.23.3 (http://cobamide2.bio.pitt.edu/computer.htm), and the Sumter genome annotation was supplemented with the Phage Evidence Collection and Annotation Network (PECAAN; https://pecaan.kbrinsgd.org/index.html). Gene function and start sites were determined using PhagesDB BLAST,

Citation Fast KM, Forrest BE, Granec GR, Larrimore JD, Morse AE, Ryan ED, Schellhammer PK, Keener TW, Sandel MW. 2019. Complete genome sequences of mycobacteriophages Candle, Schatzie, Sumter, and Waleliano. Microbiol Resour Announc 8:e00643-19. https://doi.org/10.1128/MRA.00643-19.

Editor Simon Roux, DOE Joint Genome Institute

Address correspondence to Kayla M. Fast, kfast@uwa.edu.

Received 29 May 2019
Accepted 2 July 2019
Published 25 July 2019

TABLE 1 Characteristics and accession numbers of bacteriophage genomes

Bacteriophage	GenBank accession no.	Genome size (bp)	No. of protein CDSs[a]	No. of tRNAs	G+C content (%)	Location; coordinates	Shotgun coverage (×)	Phage life cycle	Closest relative (accession no.), nucleotide identity (%), coverage (%)
Candle	MK757446	71,390	95	0	56.0	Tuscaloosa County, AL; 33.26, −87.61	564	Lytic	Send513 (JF704112), 99.26, 99.0
Schatzie	MK524521	111,345	232	1	60.8	Sumter County, AL; 32.59, −88.19	699	Lysogenic	Hughesyang (MK524504), 99.4, 89.0
Sumter	MK814754	52,656	90	0	63.7	Sumter County, AL; 32.36, −88.11	1,058	Lysogenic	Lockley (EU744249), 95.43, 83.0
Waleliano	MK524486	70,963	96	0	68.9	Baltimore County, MD; 39.25, −76.71	408	Lytic	BrownCNA (KT270441), 98.96, 99.0

[a] CDSs, coding sequences.

NCBI BLAST, Phamerator (https://phamerator.org/), GeneMark 3.25, Starterator (https://github.com/SEA-PHAGES/starterator), and HHpred 3.0 (3, 7–9). Detection of tRNAs was performed in ARAGORN 1.2.38 and tRNAscan-SE 2.0 (10, 11). Protein-coding gene and tRNA totals are provided in Table 1. Each phage was assigned to a cluster, as follows: Candle, R; Schatzie, J; Sumter, A1; and Waleliano, B4 (12). Cluster assignment employs dotplot analysis in Gepard 1.30 and comparison of average nucleotide identities (13). Similarity to the closest relative of each phage was determined using NCBI BLAST (Table 1) (7).

Data availability. The complete genome sequences of Candle, Schatzie, Sumter, and Waleliano are available from GenBank under the accession numbers MK757446, MK524521, MK814754, and MK524486, respectively. Raw Illumina reads for Candle, Schatzie, Sumter, and Waleliano are available on NCBI's Sequence Read Archive under accession numbers SRX5736299, SRX5736298, SRR8956689, and SRX5736300, respectively.

ACKNOWLEDGMENTS

This project was supported by the Howard Hughes Medical Institute Science Education Alliance–Phage Hunters Advancing Genomics and Evolutionary Science program (https://www.seaphages.org/) and the University of West Alabama 2016-2019 Phage Hunters.

We thank Daniel A. Russell and Rebecca A. Garlena for conducting genome sequencing, quality control, and assembly at the University of Pittsburgh. We thank Denise Monti with the High-Resolution Imaging Facility at the University of Alabama at Birmingham and Kimberly Lackey with the Optical Analysis Facility at the University of Alabama for assisting with transmission electron microscopy.

REFERENCES

1. Hatfull GF. 2008. Bacteriophage genomics. Curr Opin Microbiol 11:447–453. https://doi.org/10.1016/j.mib.2008.09.004.
2. Poxleitner M, Pope W, Jacobs-Sera D, Sivanathan V, Hatfull G. 2018. Phage discovery guide. Howard Hughes Medical Institute, Chevy Chase, MD.
3. Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
4. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. https://doi.org/10.1038/nature03959.
5. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res 8:195–202. https://doi.org/10.1101/gr.8.3.195.
6. Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes. Methods Mol Biol 1681:109–125. https://doi.org/10.1007/978-1-4939-7343-9_9.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
8. Borodovsky M, McIninch J. 1993. Recognition of genes in DNA sequence with ambiguities. Biosystems 30:161–171. https://doi.org/10.1016/0303-2647(93)90068-N.
9. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408.
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. Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: search and contextual analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
12. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko C, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010. Comparative genomic analysis of sixty mycobacteriophage genomes: Genome clustering, gene acquisition and gene size. J Mol Biol 397:119–143. https://doi.org/10.1016/j.jmb.2010.01.011.
13. Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23:1026–1028. https://doi.org/10.1093/bioinformatics/btm039.