



Genome Sequence and Annotation of Bacteriophage Tokki Isolated on *Arthrobacter globiformis* B-2979

Alicia Aguilera Marti,^a Nicholas D. Bullock,^a Samuel B. Everett,^a Tonya C. Bates,^a  Ellen M. Wisner^a

^aDepartment of Biological Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina

ABSTRACT *Arthrobacter* phage Tokki is a siphovirus isolated from soil in River Falls, Wisconsin. The genome contains 57,652-bp encoding 98 proteins, of which 23 were assigned a function. Tokki's genome structure and content is typical of other AU2 subcluster phages, except for the lack of an identifiable acetyltransferase.

Phages can be used to address increasing antibiotic resistance (1). Although Tokki is not currently a candidate, it expands the collection of annotated bacteriophages informing phage therapy efforts. Phage Tokki was isolated by students participating in the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program (<https://seaphages.org/>) at the University of Wisconsin-River Falls using standard procedures (2). Briefly, moist soil from River Falls, Wisconsin (44.850147N, 92.624343W) was washed with peptone-yeast extract-calcium (PYCa) medium, filtered (0.22- μ m pore size), and inoculated with *Arthrobacter globiformis* strain B-2979 with shaking at 28°C. After 2 days, the culture was filtered, diluted, and purified through two rounds of plating in top agar with *A. globiformis* at 28°C. After 24 h, Tokki exhibits clear plaques ~1 mm in diameter. Negative-stain transmission electron microscopy revealed Tokki to be a siphovirus with a capsid diameter of 59.6 nm and a tail length of 180.9 nm ($n = 1$).

DNA was isolated from phage lysate with the Promega Wizard DNA cleanup kit, prepared for sequencing using the NEB Ultra II kit, and sequenced using an Illumina MiSeq (MiSeq reagent kit v3), yielding ~624,995-bp 150 single-end reads that provided ~54-fold genome coverage. Untrimmed reads were assembled using Newbler version 2.9 (3) and finished with Consed version 29 (4, 5). Tokki has a 57,652-bp linear genome with 3' single-stranded overhangs (CGCCGCCC) and a GC content of 50.3%. Tokki is classified as a subcluster AU2 bacteriophage using previously described criteria (6, 7).

The genome was autoannotated using Glimmer 3.02 (8), GeneMark 2.5p (9), and Aragorn version 1.1 (10) embedded in DNA Master v2700 (<http://cobamide2.bio.pitt.edu>). Starterator v403 (<http://phages.wustl.edu/starterator/>), Phamerator 459 (11), and BLASTp 2.11.0 (12) were used to refine start sites. BLASTp 2.11.0 (12), HHPRED 20200717 (13), SOSUI (14), and TMHMM 2.0 (15) were used to assign gene function. While no tRNAs were identified, 98 rightward transcribed genes were identified, of which 23 were assigned a function. PECAAN 20210219 (<http://discover.kbrinsgd.org>) was used to review the annotation. All software programs used default settings.

Currently, Tokki is the largest phage in the AU2 subcluster (6) with the highest GC content (50.3%). Tokki's protein-coding genes have predicted functions associated with DNA packaging and replication (genes 2, 7, 62, 79, 83, 86, 87, and 95 to 97), structure (genes 12, 14, 16, 17, 19, 22, 23, 26 to 28, 29, and 31), and lysis (gene 3). Tokki contains unique features found in other AU2 phages, including the predicted fusion of the major capsid protein, scaffold protein and capsid maturation protein genes (gene 14); two genes are predicted that encode major tail proteins (genes 17 and 19); and the predicted tape measure protein gene (gene 27) is located in between two minor tail

Editor Kenneth M. Stedman, Portland State University

Copyright © 2022 Aguilera Marti et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ellen M. Wisner, EWisner@unc.edu.

The authors declare no conflict of interest.

Received 11 August 2022

Accepted 8 September 2022

Published 19 September 2022

proteins. Interestingly, Tokki does not appear to contain acetyltransferase, which is present in all other AU2 phages.

Accession numbers. Tokki can be accessed in GenBank with accession number [OK310503.1](https://www.ncbi.nlm.nih.gov/nuclseq/OK310503.1) and Sequence Read Archive (SRA) number [SRX14485097](https://www.ncbi.nlm.nih.gov/sra/SRX14485097).

ACKNOWLEDGMENTS

This study was supported by the Department of Biological Sciences at the University of North Carolina at Charlotte and the Howard Hughes Medical Institute Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program.

We thank Karen Klyczek and Makayla Mobeck, the instructor and student, respectively, at the University of Wisconsin-River Falls that discovered and isolated phage Tokki as part of the SEA-PHAGES project. Thanks to Dan Russell and Rebecca Garcia at the University of Pittsburgh for genome sequencing, assembly, and finishing. Thanks also to Sharon Bullock, Michelle Pass, and the annotation team from the University of North Carolina at Charlotte. We also thank Sara Tolsma from Northwestern College for feedback on the annotation.

REFERENCES

- Romero-Calle D, Guimarães Benevides R, Góes-Neto A, Billington C. 2019. Bacteriophages as alternatives to antibiotics in clinical care. *Antibiotics* 8: 138. <https://doi.org/10.3390/antibiotics8030138>.
- Poxleitner M, Pope W, Jacobs-Sera D, Sivanathan V, Hatfull GF. 2018. HHMI SEA-PHAGES phage discovery guide. <https://seaphagesphagediscoveryguide.helpdocsonline.com/home>.
- Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95:315–327. <https://doi.org/10.1016/j.ygeno.2010.03.001>.
- 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>.
- 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.
- Russell DA, Hatfull G. 2017. PhagesDB: the actinobacteriophage database. *Bioinformatics* 33:784–786. <https://doi.org/10.1093/bioinformatics/btw711>.
- Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, Russell DA, Warner MH, Hatfull GF, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES). 2017. Bacteriophages of *Gordonia* spp. display a spectrum of diversity and genetic relationships. *mBio* 8:e01069-17. <https://doi.org/10.1128/mBio.01069-17>.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
- Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 33:W451–W454. <https://doi.org/10.1093/nar/gki487>.
- 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>.
- Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* 12:395. <https://doi.org/10.1186/1471-2105-12-395>.
- 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](https://doi.org/10.1016/S0022-2836(05)80360-2).
- 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>.
- Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14:378–379. <https://doi.org/10.1093/bioinformatics/14.4.378>.
- Sonnhammer EL, Von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182.