Complete Annotated Genome Sequences of Two Novel Microbacteriophages, Gingerbug and HerculesXL, from Western Oregon, USA

Matthew R. Fisher,a Chyanna G. Blackburn,a Hope T. Poet,a Rebecca Meisnera

aBiology Department, Oregon Coast Community College, Newport, Oregon, USA

ABSTRACT  Bacteriophages Gingerbug and HerculesXL are siphoviruses that were isolated from soil in western Oregon, USA, using the actinobacterium Microbacterium foliorum. The genomes of Gingerbug and HerculesXL are similar in length and, based on gene content similarity to actinobacteriophages, were assigned to phage clusters GF and EA11, respectively.

Discovering new bacteriophages increases our understanding of their diversity, evolution, and interactions with hosts (1). Here, we report the complete genome sequences of two novel phages, Gingerbug and HerculesXL, that infect the Gram-positive bacterium Microbacterium foliorum NRRL B-24224.

Both phages were isolated using standard protocols (2). Soil samples (Table 1) were washed in peptone-yeast extract-calcium (PYCa) liquid medium. The wash was collected by centrifugation and filtration (0.22-μm pores), and the filtrate was inoculated with M. foliorum and cultured at 30°C for 2 to 5 days with shaking at 150 rpm. The culture was then filtered and plated in PYCa top agar with M. foliorum, and plates were incubated for 2 to 3 days at 30°C, yielding phages Gingerbug and HerculesXL. Both phages were purified with at least three rounds of plating. Negative-staining transmission electron microscopy revealed both phages to be siphoviruses (Fig. 1).

For each phage, double-stranded DNA was extracted from a high-titer lysate using the Promega Wizard DNA cleanup kit, prepared for sequencing using the NEBNext Ultra II FS kit, and sequenced using an Illumina MiSeq sequencer (v3 reagents). For Gingerbug, this produced 375,027 single-end 150-bp reads, which represented 1,343× average genome coverage. For HerculesXL, sequencing produced 515,560 single-end 150-bp reads, which represented 1,853× average genome coverage. For both phages, assembly was performed using Newbler v2.9 (3) and checked for completeness and characteristics of genome termini using Consed v29 (4). Both phages have circularly permuted genomes. Genome lengths and G+C contents are provided in Table 1.

We determined the open reading frames in each phage genome by running autoannotation in DNA Master v5.23.6 (http://cobamide2.bio.pitt.edu), which relies on Glimmer v3.02 (5) and GeneMark v2.5 (6). We then conducted manual review to revise the start sites and determine putative gene functions using BLAST (7) searches against the NCBI nonredundant database and the Actinobacteriophage Database, and HHpred (8) searches against the PDB

| Phage       | Genome size (bp) | G+C content (%) | No. of genes with putative functions | Total no. of genes | Cluster assignment | Soil sampling site coordinates          |
|-------------|------------------|-----------------|-------------------------------------|--------------------|--------------------|------------------------------------------|
| Gingerbug   | 39,721           | 69.3            | 32                                  | 69                 | GF                 | 44.60271N, 124.04604W                    |
| HerculesXL  | 39,410           | 63.9            | 25                                  | 60                 | EA (subcluster 11) | 44.55003N, 123.26578W                    |

Copyright © 2022 Fisher et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Matthew R. Fisher, matthew.fisher@oregoncoast.edu.

The authors declare no conflict of interest.

Received 1 September 2022
Accepted 5 October 2022
Published 26 October 2022
mmCIF70, NCBI Conserved Domain, Pfam-A, and UniProt/Swiss-Prot databases, as well as Phamerator v3.0 (9), SOSUI (10), DeepTMHMM (11), and Starterator (http://phages.wustl.edu/starterator). We searched for tRNA genes using ARAGON v1.2.38 (12) but found none. All software applications were run using default parameters.

The genome of Gingerbug contains 69 protein-coding genes, of which 32 were assigned a putative function. Seven Gingerbug genes have no homologs in the Actinobacteriophage Database (13). The genome of HerculesXL includes 60 protein-coding genes, of which 4 have no actinobacteriophage homologs and 25 were assigned a putative function (Table 1). The left halves of both genomes contain rightward-transcribed structural, assembly, and lysis genes, including a tail assembly chaperone gene with a -1 programmed translational frameshift. DNA metabolism genes are distributed across the right arms of both genomes, which contain both rightward- and leftward-transcribed genes. Based on gene content similarity of at least 35% to phages in the Actinobacteriophage Database, Gingerbug and HerculesXL were assigned to phage clusters GF and EA11, respectively (14). Neither immunity repressor nor integrase functions could be identified for either phage, suggesting that they are lytic phages.

**Data availability.** The genome sequence for Gingerbug is available in GenBank with accession no. ON970592 and Sequence Read Archive (SRA) accession no. SRR20748624. The genome sequence for HerculesXL is available in GenBank with accession no. OP068330 and SRA accession no. SRR20748623.

**ACKNOWLEDGMENTS**

We are grateful to the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program at the Howard Hughes Medical Institute, especially Viknesh Sivanathan, for material, logistical, and intellectual support. We appreciate the Pittsburgh Bacteriophage Institute providing genomic sequencing. Finally, we are indebted to the SEA-PHAGES program staff and community for advice and for reviewing the manuscript and our draft annotation.

Our laboratory work was made possible, in part, by a grant from the Siletz Tribal Charitable Contribution Fund.

**REFERENCES**

1. Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, Bhalia SS, Blumer LS, Bollivar DW, Bonilla JA, Butela KA, Coomans RJ, Cresawn SG, D’Elia T, Diaz A, Divens AM, Edgington NP, Frederic GD, Gainey MD, Garlena RA, Grant KW, Gurney SMR, Hendrickson HL, Hughes LE, Kenna MA, Klyczek KK, Kotturi H, Mavrich TN, McKinney AL, Merkhofer EC, Parker JM, Molloy SD, Monti DL, Pape-Zambito DA, Pollenz RS, Pope WH, Reyna NS, Rinehart CA, Russell DA, Shaffer CD, Sivanathan V, Stoner TH, Stuey J, Sunnen CN, Tolsma SS, Tsourkas PK, Wallen JR, Ware VC, Warner MH, Washington JM, Westover KM, White‐fleit-Smith JL, Wiersma-Koch H, Williams DC, Zack KM, Hatfull GF. 2020. Genomic diversity of bacteriophages infecting *Microbacterium* spp. PLoS One 15:e0234636. https://doi.org/10.1371/journal.pone.0234636.

2. Poxleitner M, Pope W, Jacobs-Sera D, Sivanathan V, Hatfull G. 2018. Phage discovery guide. Howard Hughes Medical Institute, Chevy Chase, MD. https://seaphagesphagediscoveryguide.helpdoconline.com/home.

3. 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.
4. 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.
5. 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.
6. 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.
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. 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.
9. Cresawn SG, Bogel M, Day N, Jacobs-Seira 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.
10. 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.
11. Hallgren J, Tsigigos KD, Pedersen MD, Armenteros JJA, Marcotilli P, Nielsen H, Krogan A, Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. https://doi.org/10.1101/2022.04.08.487609.
12. 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.
13. Russell DA, Hatfull G. 2017. PhagesDB: the Actinobacteriophage Database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
14. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Seira D, Montgomery MT, Russell DA, Warner MH, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES), Hatfull GF. 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.