Complete genome sequences of seven *Microbacterium foliorum* phages Albedo, Kenzers, Swervy, Cranjis, JaimeB, Fullmetal, and Stormbreaker

Keyshla Valentin Caban,[1] Elizabeth Kalesnik,[1] Kaitlyn A. Green,[1] Christopher J. Negro,[1] Ulises Nunez Rodriguez,[1] Milan C. Peele,[1] Casandra T. Nguyen,[1] Sydney Cahill,[1] Keara Dougherty,[1] Melissa Logue,[1] Star Hargraves,[1] Hannah Radziak,[1] Luke Willette,[1] Esther Ogunyinka,[1] Davia C. Campbell,[1] Oluwatomiwa Adebamiro,[1] Cecelia Schmeltzer,[1] Jonathan Onimus,[1] Hannah A. Asaka,[1] Wilfred Bangura,[1] Christina M. Shimp,[1] Ameera Alade,[1] Daekwon M. Sequira,[1] Tommy Jimenez,[1] Neumann University Phage Discovery Group, Sarah J. Swerdlow,[2] Melinda K. Harrison,[3] Patricia C. Fallest-Strobl,[1] Matthew D. Mastropaolo[1]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Seven bacteriophages were isolated from soil in Pennsylvania and Wisconsin using the host *Microbacterium foliorum*. These bacteriophages range in the number of predicted genes encoded, from 25 to 91, and are distributed across actinobacteriophage clusters EB, EC, EE, and EK.

Bacteriophages are incredibly abundant and genetically diverse. To expand our knowledge of bacteriophage evolution and diversity, we report here the characteristics of seven bacteriophages newly isolated using *Microbacterium foliorum* NRRL B-24224 (1, 2).

All seven bacteriophages were isolated from soil in Pennsylvania and Wisconsin using standard methods as previously described (Table 1) (3, 4). These soil samples were incubated in peptone-yeast extract-calcium (PYCa) liquid medium for 2 hours at 30°C with shaking to suspend phage particles. The suspension was then filtered through a 0.22-µm filter. The filtrate was either directly plated in PYCa soft agar containing *M. foliorum* or "enriched" by inoculation with *M. foliorum* and incubation at 30°C for 2–3 days before being filtered and plated (Table 1), yielding phages Albedo, Kenzers, Swervy, Cranjis, JaimeB, Fullmetal, and Stormbreaker. All phages were purified through three rounds of plating. All plates were incubated at 30°C for 24–48 hours.

The Wizard DNA Cleanup Kit (Promega) was used to extract genomic DNA from phage lysates, as previously described (4). Some lysates were concentrated using $ZnCl_2$ precipitation prior to genomic DNA extraction (6). The genomic DNA libraries were prepared using a NEBNext Ultra II FS Kit (New England BioLabs) followed by sequencing using Illumina MiSeq (v3 reagents), yielding at least 40,000 150-base single-end reads (Table 1). Raw reads were assembled and then checked for completeness using Newbler v2.9 (7) and Consed v29 (8), respectively (9). Sequencing results and genome characteristics of each bacteriophage are listed in Table 1.

The genomes were autoannotated using DNA Master v5.23.6 (http://cobamide2.bio.pitt.edu), Glimmer v3.02b (10), GeneMark v4.28 (11) and were refined using PECAAN v20221109 (https://pecaan.kbrinsgd.org/index.html), Starterator v462 (https://github.com/SEA-PHAGES/starterator), and Phamerator v539 (12). Transmembrane helices were predicted using SOSUI v1.11 (13), TOPCONS v2.0 (14), TMHMM v2.0 (15), and DeepTMHMM v1.0.24 (16). tRNAs were predicted using ARAGORN v1.2.41 (17) and tRNAscanSE v2.0 (18). Putative functions for other predicted genes were made using

TABLE 1 Bacteriophage, plaque morphology, and genomic characteristics

Phage name	Soil sample collection site	Isolation method	Plaque morphology	Plaque Size[a] (mm)	Approx. shotgun coverage (fold)	No. of 150-bp single-end reads	Genome length (bp)	Genome end characteristic	G + C content (%)	No. of ORFs[b]	No. of tRNAs	Cluster[c]
Albedo	Hudson, WI, 44.984533 N, 92.7545 W	Enriched	Clear with halo[d]	0.1–1	83	862,002	41,813	3′ single-stranded overhang 5′-TCTCCCGGCA-3′	66.6	71	1 (Gln)	EB
Kenzers	Greenville, PA, 41.4124 N, 80.3813 W	Enriched	Clear	1	3,293	960,944	41,261	3′ single-stranded overhang 5′-TCTCCCGGCA-3′	66.8	70	1 (Gln)	EB
Swervy	Aston, PA, 39.8657 N, 75.4279 W	Direct	Turbid	0.5–1	325	93,814	41,510	3′ single-stranded overhang 5′-TCTCCCGGCA-3′	66.7	71	1 (Asn)	EB
Cranjis	Upper Chichester, PA, 39.856232 N, 75.443149 W	Direct	Turbid	3.5–4	99	40,843	53,222	Circularly permutated	68.9	91	0	EC
JaimeB	Aston, PA, 39.875331 N, 75.440021 W	Direct	Clear	3–4	14,635	1.8 million	17,445	3′ single-stranded overhang 5′-CCCGCCCCA-3′	68.7	25	0	EE
Stormbreaker	Aston, PA, 39.5215 N, 75.260806 W	Direct	Clear	1	1,073	74,440	54,050	Circularly permutated	60	54	0	EK[e]
Fullmetal	Aston, PA, 39.876667 N, 75.441667 W	Direct	Clear	1–1.5	197	410,171	54,438	Circularly permutated	59.8	55	0	EK[f]

[a]Plaque size is based on the measurements of three plaques.
[b]ORFs, open reading frames.
[c]Clusters were identified using sequence similarities to other *Microbacterium* phage (5).
[d]Indicates a clear middle of the plaque with a diffuse or cloudy edge.
[e]Subcluster EK2.
[f]Subcluster EK1.
[g]"ND" indicates that the TEM was not performed.
[h]Isolation methods are described in the Phage Discovery Guide (3, 4).

HHPRED v3.2 (against the PDB_mmCIF70, NCBI_Conserved_Domains, Pfam-A, and UniProt-SwissProt databases) (19) and BlastP v2.10.0 (against the PhagesDB and NCBI nonredundant databases) (20). All annotations were performed with default parameters.

Phages were assigned to clusters based on gene content similarity (GCS) of at least 35% to sequenced genomes in the Acinobacteriophage database (https://phagesdb.org/) using the GCS tool at phagesDB (5, 21). All seven phages reported here are consistent with features previously described for their respective clusters; the EB cluster phages, Albedo, Kenzers, and Swervy encode for <3 tRNAs; the EC cluster phage Cranjis has all its genes transcribed rightward; the EE cluster phage JaimeB shares all 25 predicted genes including a capsid maturation and protease fusion protein with the other EE cluster members; the EK cluster phages Stormbreaker and Fullmetal have the first ~30 predicted genes transcribed leftward and all the remaining genes transcribed rightward, and they also encode for the largest actinobacteriophage gene product, over 4,400 amino acids (1, 22).

AUTHOR AFFILIATIONS

[1]Sciences, Neumann University, Aston, Pennsylvania, USA
[2]Biological Sciences, University of Pittsburgh Greensburg, Greensburg, Pennsylvania, USA
[3]Sciences, Cabrini University, Radnor, Pennsylvania, USA

AUTHOR ORCIDs

Melinda K. Harrison http://orcid.org/0000-0003-2047-7087
Matthew D. Mastropaolo http://orcid.org/0000-0002-8148-4742

AUTHOR CONTRIBUTIONS

Keyshla Valentin Caban, Investigation, Validation | Elizabeth Kalesnik, Investigation, Validation | Kaitlyn A. Green, Investigation, Validation | Christopher J. Negro, Investigation, Validation | Ulises Nunez Rodriguez, Investigation, Validation | Milan C. Peele, Investigation, Validation | Casandra T. Nguyen, Investigation, Validation | Sydney Cahill, Investigation, Validation | Keara Dougherty, Investigation, Validation | Melissa Logue, Investigation, Validation | Star Hargraves, Investigation, Validation | Hannah Radziak, Investigation, Validation | Luke Willette, Investigation, Validation | Esther Ogunyinka, Investigation, Validation | Davia C. Campbell, Investigation, Validation | Oluwatomiwa Adebamiro, Investigation, Validation | Cecelia Schmeltzer, Investigation, Validation | Jonathan Onimus, Investigation, Validation | Hannah A. Asaka, Investigation, Validation | Wilfred Bangura, Investigation, Validation | Christina M. Shimp, Investigation, Validation | Ameera Alade, Investigation, Validation | Daekwon M. Sequira, Investigation, Validation | Tommy Jimenez, Investigation, Validation | Sarah J. Swerdlow, Validation, Writing – review and editing | Melinda K. Harrison, Data curation, Investigation, Validation, Writing – review and editing | Patricia C. Fallest-Strobl, Data curation, Investigation, Validation | Matthew D. Mastropaolo, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing.

DATA AVAILABILITY

All genomes, Albedo, Kenzers, Swervy, Cranjis, JaimeB, Fullmetal, and Stormbreaker are available at GenBank with Accession No. OR475283, OP172875, MZ747513,

OP297543, OR195050, OP297538, MT657334 and the Sequence Read Archive (SRA) No. SRX22868877, SRX14483228, SRX14485092, SRX22853654, SRX22853656, SRX22853655, SRX22853658, respectively.

REFERENCES

1. Mastropaolo MD, Fallest-Strobl PC, Sequira DM, Campbell DD, Negro CJ, Tuang KS, Sigmund-Hamre IM, Womack CL, Mansbridge T, Metzler AL, Sasher EG, Collins C, Crowley NK, Dower VR, Bates M, Bjorkelo C, Johnson H, Salvitti LR, Neumann University Phage Discovery Group. 2023. Genome sequences of five *Microbacterium foliorum* phages, GaeCeo, NeumannU, Eightball, Chivey, and Hiddenleaf. Microbiol Resour Announc 12:e0110622. https://doi.org/10.1128/mra.01106-22

2. Hatfull GF. 2020. Actinobacteriophages: genomics, dynamics, and applications. Annu Rev Virol 7:37–61. https://doi.org/10.1146/annurev-virology-122019-070009

3. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, et al. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. mBio 5:e01051-13. https://doi.org/10.1128/mBio.01051-13

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

5. Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711

6. Santos MA. 1991. An improved method for the small scale preparation of bacteriophage DNA based on phage precipitation by zinc chloride. Nucleic Acids Res 19:5442. https://doi.org/10.1093/nar/19.19.5442

7. 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

8. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics 29:2936–2937. https://doi.org/10.1093/bioinformatics/btt515

9. Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes. Methods Mol Bio 1681:109–125. https://doi.org/10.1007/978-1-4939-7343-9

10. 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

11. 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/gki487.13

12. 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

13. 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

14. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res 43:W401–W407. https://doi.org/10.1093/nar/gkv485

15. Krogh A, Larsson B, von Heijne 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

16. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh 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

17. 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

18. Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413

19. 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

20. 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

21. 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

22. Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, Bhalla SS, Blumer LS, Bollivar DW, Bonilla JA, Butela KA, et al. 2020. Genomic diversity of bacteriophages infecting *Microbacterium* spp. PLoS One 15:e0234636. https://doi.org/10.1371/journal.pone.0234636