AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of the Cluster P Mycobacteriophage Phegasus

Abigail A. Howell,[a,b] Cyril J. Versoza,[a,c] Gabriella Cerna,[a,b,d] Tyler Johnston,[d] Shriya Kakde,[a] Keith Karuku,[a] Maria Kowal,[a] Jasmine Monahan,[a] Jillian Murray,[a,d] Teresa Nguyen,[a] Aurely Sanchez Carreon,[a,b,d] Elizabeth Song,[e] Abigail Streiff,[d] Blake Su,[a,f] Faith Youkhana,[d] Saige Munig,[a] Zeel Patel,[a] Minerva So,[a] Makena Sy,[a] Sarah Weiss,[a] Yang Zhou,[e] Susanne P. Pfeifer[a,c,g]

[a]School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[b]Biodesign Institute, Arizona State University, Tempe, Arizona, USA
[c]Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, USA
[d]School of Molecular Sciences, Arizona State University, Tempe, Arizona, USA
[e]Division of Biology and Medicine, Brown University, Providence, Rhode Island, USA
[f]School of Politics and Global Studies, Arizona State University, Tempe, Arizona, USA
[g]Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

Abigail A. Howell and Cyril J. Versoza contributed equally to this article. Author order was determined alphabetically.

ABSTRACT We characterized the complete genome of the cluster P mycobacteriophage Phegasus. Its 47.5-kb genome contains 81 protein-coding genes, 36 of which could be assigned a putative function. Phegasus is most closely related to two subcluster P1 bacteriophages, Mangethe and Majeke, with an average nucleotide identity of 99.63% each.

A diverse range of bacteriophages is known to infect *Mycobacterium smegmatis* (1). As part of the Howard Hughes Medical Institute Science Education Alliance—Phage Hunters Advancing Genomics and Evolutionary Science (HHMI SEA-PHAGES) program, we characterized the complete genome of Phegasus, a putatively temperate cluster P, subcluster P1 mycobacteriophage.

Phegasus was obtained from a soil sample collected from the manure area of a horse barn at the Guilford Riding School (Guilford, CT; 41.3029 N, 72.6537 W) through enriched isolation, purification, and amplification in *Mycobacterium smegmatis* mc^{2}155, following the procedures outlined in the SEA-PHAGES Discovery Guide (https://seaphagesphagediscoveryguide.helpdocsonline.com/home). A dual-indexed sequencing library was prepared from genomic DNA using the NEBNext Ultra II FS kit and sequenced on an Illumina MiSeq instrument (coverage: >900×). Following Russell (2), Newbler v.2.9 was used to *de novo* assemble the 307,831 single-end (150-bp) reads into a full-length genome sequence, with a 12-base 3′ sticky overhang. The 47,578-bp genome exhibits a GC content of 67.4%. The completeness, accuracy, and genomic termini were checked using Consed v.29.0 (3). All software was executed using default settings.

Genome annotation followed the HHMI SEA-PHAGES Bioinformatics Guide (https://seaphagesbioinformatics.helpdocsonline.com/home), using GLIMMER v.3.0.2 (4) and GeneMark v.2.5 (5) embedded within DNA Master v.5.23.6 to identify open reading frames. Eighty-one protein-coding genes were predicted in the genome (gene density: 1.70 genes/kb), of which 36 could be assigned a putative function using NCBI BLAST (6) and HHpred (7), as well as information on synteny obtained using Phamerator (8). Of the remaining genes, five were classified as membrane proteins using TMHMM v.2.0 (9) and SOSUI v.1.11 (10). The left arm of the genome encodes several well-conserved structural and assembly proteins (including small and large terminase subunits, a portal protein, capsid maturation protease, a scaffolding protein, a major capsid protein, both a head-to-tail adapter and stopper, a tail terminator, a major tail protein, two tail assembly chaperones, a tape measure

Editor Kenneth M. Stedman, Portland State University

Address correspondence to Susanne P. Pfeifer, susanne.pfeifer@asu.edu.

The authors declare no conflict of interest.

Received 9 June 2022
Accepted 21 July 2022

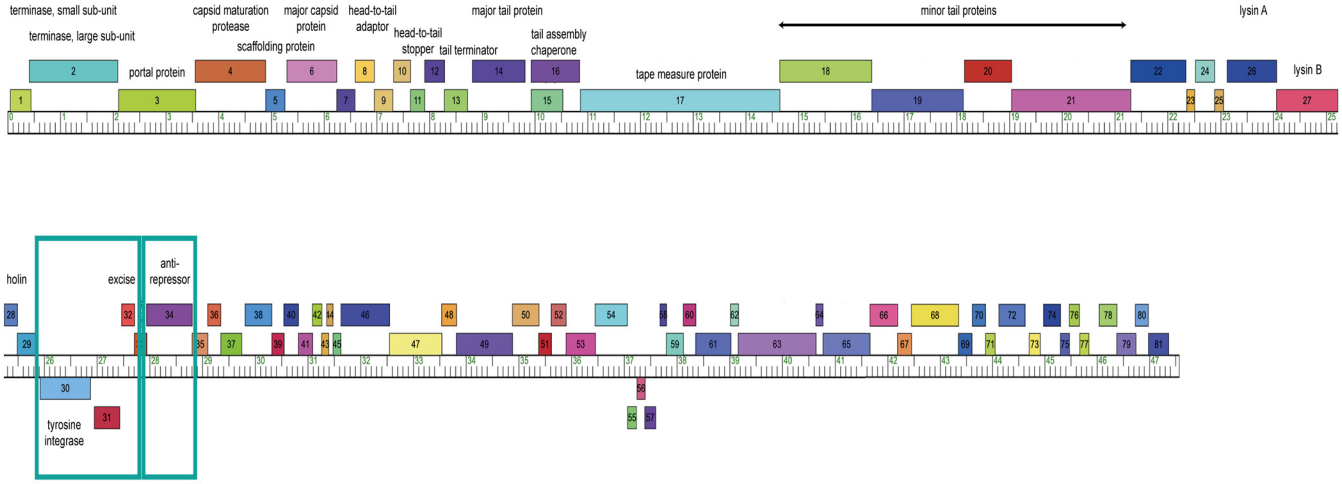

FIG 1 Genome of the cluster P mycobacteriophage Phegasus. Protein-coding genes on the forward or reverse strands with their putative functional assignments (if available) are displayed above or below the ruler, respectively. The integration-dependent immunity system (genes 30 to 32 and 34) is indicated by teal-colored boxes. ssDNA, single-stranded DNA.

protein, and four minor tail proteins). Following the structural proteins is a lysin cassette, comprised of lysin A and lysin B, responsible for the cleavage of the host cell wall during the final stages of the lytic cycle. The right arm of the genome encodes nonstructural genes, including an integration-dependent immunity system (genes 30 to 32 and 34) that governs the transition from the lysogenic to lytic state (Fig. 1). A partial tRNA (located at positions 26972 to 27076) was identified using tRNAscan-SE v.2.0 (Infernal score, 12.6) (11), which may represent either the remnants of a full-length tRNA or part of a tRNA that is assembled after integration into the host genome.

Multiple sequence alignments were generated using MAFFT v.7 (12), which demonstrated that Phegasus is most closely related to two subcluster P1 bacteriophages, Mangethe (GenBank accession number MK016499) and Majeke (MF472894), collected at the University of KwaZulu-Natal in South Africa, with an average nucleotide identity of 99.63% each.

Data availability. The whole-genome sequencing data are available at NCBI's Sequence Read Archive (accession number SRR19912416 and BioProject accession number PRJNA488469). The annotated genome assembly is available at NCBI GenBank under accession number ON637760.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation CAREER grant to S.P.P. (DEB-2045343), Howard Hughes Medical Institute's SEA-PHAGES program, and Arizona State University's School of Life Sciences. Bacteriophage isolation was performed at Brown University (Providence, RI); library preparation, sequencing, and *de novo* assembly were performed at the University of Pittsburgh (Pittsburgh, PA); and genome annotations and comparative analyses were performed at Arizona State University (Tempe, AZ).

We are grateful to Suhail Ghafoor for IT support, Daniel A. Russell and Rebecca A. Garlena for *de novo* assembly, as well as Billy Biederman, Graham Hatfull, Deborah Jacobs-Sera, and Vic Sivanathan for training and continued support in the SEA-PHAGES program.

REFERENCES

1. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education, Mycobacterial Genetics Course. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4:e06416. https://doi.org/10.7554/eLife.06416.
2. 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.

3. 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.
4. 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.
5. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res 26:1107–1115. https://doi.org/10.1093/nar/26.4.1107.
6. 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.
7. 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.
8. 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.
9. 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.
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. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.
12. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.