



Complete Genome Sequence of the *Pantoea* Phage AH07

Sam J. Roth,^a Greg P. Krukonis,^b  Véronique A. Delesalle^a

^aDepartment of Biology, Gettysburg College, Gettysburg, Pennsylvania, USA

^bDepartment of Biology, Angelo State University, San Angelo, Texas, USA

ABSTRACT Bacteriophages of the phyllosphere have not been extensively described, despite their role in bacterial communities on this plant organ. Here, we describe a temperate *Pantoea* phage, AH07, that was isolated from the leaves of horse chestnut trees. The 37,859-bp linear double-stranded DNA genome contains 58 putative genes, including an integration cassette.

As part of a broad evolutionary research program (1–5), bacteriophages from the leaves of horse chestnut trees (*Aesculus hippocastanum*; Sapindaceae) from Angel and Greyhound Meadow (Oxford, UK) were isolated in 2011 to 2014 on bacterial strains that themselves were isolated from these leaves (1). Bacterial isolates were assigned to a given genus and, if possible, species based on sequencing of approximately 800 bp of the 16S rRNA region and the top BLASTn hit associated with a sequence (E value of $<10^{-10}$) (1). Individual phages were each single plaque purified at least three times on their focal hosts and amplified by overnight culturing in 10 ml King's broth with 100 μ l of isolation bacteria (1); we refer to these as phyllophages AH01 to AH07. Here, we describe the temperate phage AH07, which was isolated on a *Pantoea* species.

The culture lysate was filtered (0.45 μ m), and DNA extraction was performed by the Koskella laboratory following the protocol for the Wizard PCR Preps DNA purification system kit (product no. 7170; Promega). Library preparation and sequencing were performed at the North Carolina State University Genomic Science Laboratory. Libraries were prepared using an Illumina TruSeq Nano DNA library preparation kit, following the manufacturer's protocol. Sequencing was done with an Illumina MiSeq platform, using a v3 single-end 150-bp flow cell. Genome assembly and annotation were performed at Gettysburg College. The 409,887 single-end 150-bp reads were aligned and assembled into one contig with 1,624 \times coverage using GS *De Novo* Assembler v2.9; all untrimmed reads were used, and contig quality was verified in Consed v29 (6, 7). Genome ends were determined with PAUSE (<https://cpt.tamu.edu/analysis-with-pause3-2016-edition>) and PhageTerm (8). The finished sequence was imported into DNA Master v5.22.22 (<http://cobamide2.bio.pitt.edu/computer.htm>) to map open reading frames. Putative genes were called based on both Glimmer v3.0 and GeneMark v2.5 algorithms (9, 10). Putative functions of gene products were predicted using BLAST v2.12 (11) and HHpred (12). For BLASTp matches, an E value of $<10^{-5}$ was required to assign function. For HHpred matches, a high probability ($>85\%$), substantial coverage ($>50\%$), and low E value ($<10^{-5}$) were required. Default settings were used with all of the aforementioned programs.

AH07 has a 37,859-bp genome with a GC content of 52.2%, 58 putative genes, including 39 genes with assigned functions, and no tRNA genes. The genome architecture of AH07 follows the common organization of *Siphoviridae* phages, with an ordered cluster of structural genes followed by a mixture of genes involved in DNA metabolism or of unknown function (13). The integration cassette (tyrosine integrase, excise, and immunity repressor) follows the structural genes, and the lysis cassette (holin, endolysin, and Rz-like lysis protein) is located at the end of the genome. A phage isolate obtained from a different leaf on the same tree in 2012 was sequenced following the

Citation Roth SJ, Krukonis GP, Delesalle VA. 2021. Complete genome sequence of the *Pantoea* phage AH07. *Microbiol Resour Announc* 10:e00819-21. <https://doi.org/10.1128/MRA.00819-21>.

Editor Kenneth M. Stedman, Portland State University

Copyright © 2021 Roth 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 Véronique A. Delesalle, delesalle@gettysburg.edu.

Received 23 August 2021

Accepted 11 October 2021

Published 4 November 2021

TABLE 1 Results of whole-genome (nucleotide) searches for AH07 using BLASTn

Database	Sample description ^a	Query coverage (%)	Identity (%)	Accession no.
Complete	<i>Pantoea agglomerans</i> strain CFSAN047153	69	89.64	CP034469.1
Complete	<i>Pantoea agglomerans</i> strain CFSAN047154	69	89.64	CP034474.1
Complete	<i>Pantoea vagans</i> strain PV989	62	92.44	CP028349.1
Complete	<i>Pantoea agglomerans</i> strain ASB05	49	89.64	CP046722.1
Restricted to phages	TPA asm: <i>Siphoviridae</i> sp. strain ct4sp3	58	93.94	BK015502.1
Restricted to phages	TPA asm: <i>Siphoviridae</i> sp. isolate ctTyS5	48	87.33	BK029822.1

^aThe complete genome of AH07 was compared to the complete nucleotide database or the same database restricted to tailed phages (taxid 10699, 10662, and 10744). For each search, the best matches are reported (defined as more than 45% query coverage). The percentage of the query covered by alignment to the database sequence and the percent identity of matched alignment are reported. TPA asm, third-party annotation assembly.

protocols described above and was determined to be identical to AH07. From whole-genome searches by BLASTn (Table 1), AH07 was found to have the greatest nucleotide similarity to segments of *Pantoea agglomerans* and *Pantoea vagans* genomes, supporting its temperate life cycle. When the BLASTn search was restricted to phages (Table 1), the best matches were to *Siphoviridae* samples, intriguingly isolated from throughout the human body (14).

Data availability. The genome sequence and associated information can be found under GenBank accession no. [MZ501270](https://ncbi.nlm.nih.gov/nucl/MZ501270), SRA accession no. [SRX11736858](https://ncbi.nlm.nih.gov/sra/SRX11736858), and BioProject accession no. [PRJNA754193](https://ncbi.nlm.nih.gov/bioproject/PRJNA754193).

ACKNOWLEDGMENTS

This research was supported by Research and Professional Development grants from Gettysburg College to V.A.D.

We thank the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program, especially Graham Hatfull, Welkin Pope, Dan Russell, and Debbie Jacobs-Sera, for providing training in genome annotation and answering all of our phage questions. We thank Britt Koskella for providing us with phage DNA to sequence and the opportunity to learn more about the phyllosphere.

REFERENCES

- Koskella B, Thompson JN, Preston GM, Buckling A. 2011. Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. *Am Nat* 177:440–451. <https://doi.org/10.1086/658991>.
- Koskella B. 2013. Phage-mediated selection on microbiota of a long-lived host. *Curr Biol* 23:1256–1260. <https://doi.org/10.1016/j.cub.2013.05.038>.
- Koskella B, Parr N. 2015. The evolution of bacterial resistance against bacteriophages in the horse chestnut phyllosphere is general across both space and time. *Philos Trans R Soc B Biol Sci* 370:20140297. <https://doi.org/10.1098/rstb.2014.0297>.
- Koskella B, Meaden S. 2013. Understanding bacteriophage specificity in natural microbial communities. *Viruses* 5:806–823. <https://doi.org/10.3390/v5030806>.
- Morella NM, Gomez AL, Wang G, Leung MS, Koskella B. 2018. The impact of bacteriophages on phyllosphere bacterial abundance and composition. *Mol Ecol* 27:2025–2038. <https://doi.org/10.1111/mec.14542>.
- Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. *Bioinformatics* 29:2936–2937. <https://doi.org/10.1093/bioinformatics/btt515>.
- Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes. *Methods Mol Biol* 1681:109–135. https://doi.org/10.1007/978-1-4939-7343-9_9.
- Garneau JR, Depardieu F, Fortier LC, Bikard D, Monot M. 2017. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci Rep* 7:8292. <https://doi.org/10.1038/s41598-017-07910-5>.
- 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>.
- 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>.
- Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biol Direct* 7:12. <https://doi.org/10.1186/1745-6150-7-12>.
- Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:244–248. <https://doi.org/10.1093/nar/gki408>.
- Hatfull GF. 2012. The secret lives of mycobacteriophages. *Adv Virus Res* 82:179–288. <https://doi.org/10.1016/B978-0-12-394621-8.00015-7>.
- Tisza MJ, Buck CB. 2021. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc Natl Acad Sci U S A* 118:e2023202118. <https://doi.org/10.1073/pnas.2023202118>.