# Complete Genome Sequence of *Rhodococcus erythropolis* Phage Shuman

Sucely Ponce Reyes,[a] Peter J. Park,[a] Daniel Kaluka,[a] Jacqueline M. Washington[a]

[a]Department of Natural Sciences, Nyack College, Nyack, New York, USA

**ABSTRACT** Shuman is a bacteriophage isolated in Nyack, New York, using *Rhodococcus erythropolis* NRRL B-1574 as a host. It is a member of cluster CA and has a genome length of 46,544 bp. Shuman contains 67 predicted protein-coding genes, 3 tRNA genes, and no transfer-messenger RNA (tmRNA) genes.

Shuman is a bacteriophage that was isolated from soil collected at the Nyack College campus in Nyack, New York, in 2017 by using *Rhodococcus erythropolis* NRRL B-1574 as a host. The phage was isolated by direct plating at 30°C on peptone-yeast extract-calcium (PYCa) medium supplemented with 0.1% dextrose and forms turbid plaques 1 to 2 mm in diameter. Briefly, soil was incubated with PYCa to obtain a soil extract which was filtered through a 0.22-$\mu$m filter, and 50 $\mu$l of this extract was added to 250 $\mu$l late-exponential-phase *Rhodococcus erythropolis* and plated with 3 ml molten PYCa top agar (0.35%). To purify, well-isolated individual plaques were picked and added to phage buffer with glycerol (10 mM Tris-HCl [pH 7.5], 10 mM MgSO$_4$, 1 mM CaCl$_2$, 68 mM NaCl, and 10% glycerol), serially diluted, and replated for three successive rounds.

DNA was isolated from a high-titer lysate of purified phage by using a Promega Wizard DNA extraction kit. A sequencing library was prepared from genomic DNA by using an NEB Ultra II FS kit with dual-indexed barcoding. It was multiplexed with 47 other phage genome libraries and run on an Illumina MiSeq instrument, yielding ~555,000 single-end 150-base reads from the Shuman sample, representing ~1,684-fold coverage of the genome. These raw reads were assembled using Newbler version 2.9 with default settings, yielding a single-phage contig (46,544 bp) which was checked for completeness, accuracy, and phage genomic termini using Consed version 29 and as previously described (1). The Shuman genome is 46,544 bp in length, with 58.6% GC content, and has defined ends with 10-base 3′ single-stranded extensions (CGGCCGT GAT). Annotation analysis was performed using the following databases and software as of July 2018 with default settings: DNA Master (http://cobamide2.bio.pitt.edu/computer.htm), PECAAN (https://pecaan.kbrinsgd.org/), Phamerator (2), Glimmer 3.0 (3), Genemark 2.5 (4), HHPRED (5), NCBI BLAST 2.7 and Conserved Domain Database at NCBI (6), ARAGORN version 1.2.38 (7), tRNA scan-SE 3.0 (8), and TMHMM 2.0 (9, 10).

Based on >90% nucleotide similarity across the entire genome, Shuman is classified as a member of cluster CA, the most common cluster containing *Rhodococcus* phages, to which 70% of the sequenced *Rhodococcus* phages belong (11). Nearly all cluster CA phages have been isolated using *Rhodococcus erythropolis* RIA 643. Cluster CA phages are members of the *Siphoviridae* family and are predicted to be temperate.

Sixty-seven protein-coding genes and 3 tRNAs were identified in the Shuman genome. Similarly to the other cluster CA phages, ~50% of the predicted protein-coding genes were assigned functions, and the tRNAs are within 2 kb of the left end of the genome (12). Four genes with no known function are predicted to form transmembrane helices. No transfer-messenger RNAs (tmRNAs) were identified. Shuman is most similar to StCroix (GenBank accession number MF324900), which was isolated in 2014

in Hudson, Wisconsin, with 98% identity over 99% coverage. The Shuman genome is arranged with the structural and assembly genes on the left end while the regulatory and replication genes are on the right end of the genome.

**Data availability.** Bacteriophage Shuman genome sequence is available at GenBank under the accession number MH316569. The raw sequences are available in the NCBI SRA database under the accession number SRR8477199.

## ACKNOWLEDGMENTS

## REFERENCES

1. Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes. p 109–125. *In*: Clokie M, Kropinski A, Lavigne R. (ed), Bacteriophages. Methods in molecular biology, vol 1681. Humana Press, New York, NY.
2. 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.
3. 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.
4. 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.
5. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.
6. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222–D226. https://doi.org/10.1093/nar/gku1221.
7. 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.
8. Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: integrating search and contextual analysis of transfer RNA Genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
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 Molecular Biology 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
10. Sonnhammer EL, von Heijne G., Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences, p 175–182. *In* Glasgow J, Littlejohn T, Major F, Lathrop R, Sankoff D, and Sensen C, (ed), Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA.
11. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J Mol Biol 397:119–143. https://doi.org/10.1016/j.jmb.2010.01.011.
12. Bonilla JA, Isern S, Findley AM, Klyczek KK, Michael SF, Saha MS, Buchser WJ, Forsyth MH, Paudel S, Gissendanner CR, Wiedemeier AMD, Alonzo FL, University of Wisconsin–River Falls SEA-PHAGES, Florida Gulf Coast University SEA-PHAGES, University of Louisiana–Monroe SEA-PHAGES, College of William & Mary SEA-PHAGES, Garlena RA, Russell DA, Pope WH, Cresawn SG, Jacobs-Sera D, Hatfull GF. 2017. Genome sequences of 19 *Rhodococcus erythropolis* cluster CA phages. Genome Announc 5:e01201-17. https://doi.org/10.1128/genomeA.01201-17.