Genome Sequences of 19 *Rhodococcus erythropolis* Cluster CA Phages

J. Alfred Bonilla,[a] Sharon Isern,[b] Ann M. Findley,[c] Karen K. Klyczek,[a]
Scott F. Michael,[b] Margaret S. Saha,[d] William J. Buchser,[d] Mark H. Forsyth,[d]
Sudip Paudel,[d] Christopher R. Gissendanner,[c] Allison M. D. Wiedemeier,[c]
Fernanda L. Alonzo,[c] University of Wisconsin–River Falls SEA-PHAGES,[a]
Florida Gulf Coast University SEA-PHAGES,[b] University of Louisiana–Monroe
SEA-PHAGES,[c] College of William & Mary SEA-PHAGES,[d] Rebecca A. Garlena,[e]
Daniel A. Russell,[e] Welkin H. Pope,[e] Steven G. Cresawn,[f]
Deborah Jacobs-Sera,[e] Graham F. Hatfull[e]

Department of Biology, University of Wisconsin–River Falls, River Falls, Wisconsin, USA[a]; Department of Biological Sciences, Florida Gulf Coast University, Fort Myers, Florida, USA[b]; Biology Program, School of Sciences, University of Louisiana–Monroe, Monroe, Louisiana, USA[c]; Department of Biology, College of William & Mary, Williamsburg, Virginia, USA[d]; Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA[e]; Department of Biology, James Madison University, Harrisonburg, Virginia, USA[f]

ABSTRACT   We report the complete genome sequences of 19 cluster CA bacteriophages isolated from environmental samples using *Rhodococcus erythropolis* as a host. All of the phages are *Siphoviridae*, have similar genome lengths (46,314 to 46,985 bp) and G+C contents (58.5 to 58.8%), and share nucleotide sequence similarity.

A large collection of sequenced mycobacteriophages reveals substantial genetic diversity and mosaic genomic architectures (1). Smaller collections of phages have been isolated on other hosts within the phylum *Actinobacteria*, including *Rhodococcus* spp. (2). Previously described *Rhodococcus* phages include DocB7, Pepy6, Pine5, Poco6, REQ1, REQ2, REQ3, and E3 isolated on *R. equi*, and RER2, RGL3, and RRH1 isolated on *R. erythropolis*, *R. globerulus*, and *R. rhodochrous*, respectively (3–7), representing nine different genome types (2).

Students in the Howard Hughes Medical Institute (HHMI) Science Education Alliance–Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program (8) isolated phages infecting *R. erythropolis* RIA 643 using environmental samples from diverse geographical locations (Table 1). Following purification and amplification, genomic DNA was isolated and sequenced using Illumina MiSeq 150-base reads and assembled with Newbler and Consed to major contigs with a 150-fold minimum coverage. Genomes were annotated using DNA Master (http://cobamide2.bio.pitt.edu/), Glimmer (9), GeneMark (10), ARAGORN (11), and tRNAscan-SE (12). Functions were assigned using BLASTp (13), HHpred (14), and Phamerator (15). The 19 genomes have similar lengths (46,314 to 46,985 bp) and G+C contents (58.5 to 58.8%), and have defined genome ends with 10-base 3′ single-stranded extensions (5′-CGGCCGTGAT). All share nucleotide sequence similarity (>93% pairwise DNA identity) with each other and with phages RER2 and RGL3 (7), and are grouped in cluster CA. CosmicSans, Lillie, Rhodalysa, and TWAMP were isolated from distinct but geographically similar locations and have similar genomes, differing by 2 to 6 nucleotides.

We identified 64 to 67 predicted protein-coding genes in each genome, of which ~50% were ascribed putative functions, as well as 2 to 3 tRNA genes located approximately 1.5 kbp from the left genome end. The genome architectures are reminiscent of the cluster A mycobacteriophages such as L5 (16), with the virion structure and assembly genes transcribed rightward in the left arms and regulatory and replication genes transcribed leftward in the right arms. All of the cluster CA *Rhodococcus* phages

TABLE 1 Cluster CA phages of *Rhodococcus erythropolis*

Phage	GenBank accession no.	Genome size (bp)	G+G content (%)	No. of PCGs[a]	Location of isolation
Alatin	MF324905	46,673	58.7	65	Hudson, WI
Angryorchard	KY549153	46,597	58.8	64	Somerville, TN
AppleCloud	MF324903	46,389	58.7	65	Hudson, WI
BobbyDazzler	KY549154	46,641	58.8	65	Monroe, LA
Bonanza	MF537628	46,932	58.8	66	Keithville, LA
CosmicSans	KT372002	46,596	58.5	66	Williamsburg, VA
Harlequin	KX611788	46,383	58.8	66	Monroe, LA
Hiro	MF324898	46,854	58.7	65	Boston, MA
Jester	MF373842	46,314	58.7	66	Ouachita Parish, LA
Krishelle	MF324902	46,985	58.5	67	Hudson, WI
Lillie	KT990218	46,596	58.6	66	Williamsburg, WI
Naiad	MF324901	46,619	58.6	65	Hudson, WI
Natosaleda	KX550082	46,527	58.6	65	Boston, MA
Partridge	KX712237	46,962	58.8	66	Monroe, LA
RexFury	MF324904	46,627	58.6	65	River Falls, WI
Rhodalysa	KT375356	46,596	58.5	66	Williamsburg, VA
StCroix	MF324900	46,619	58.6	65	Hudson, WI
TWAMP	KT959213	46,596	58.5	66	Williamsburg, VA
Yogi	KX712236	46,930	58.8	66	Richland Parish, LA

[a]PCGs = protein-coding genes.

are predicted to be temperate and encode putative immunity repressors (e.g., Alatin gp57) and serine integrases (e.g., Alatin gp29) close to the genome center; the chromosomal integration site is not known. Stable lysogens have been isolated for several of the phages, and all of those tested are homoimmune.

Although these *Rhodococcus* phages do not share extensive nucleotide sequence similarity to the cluster A mycobacteriophages, over 40% of the predicted gene products have amino acid sequence similarity to cluster A gene products. They likely also share a regulatory system with cluster A phages in which the immunity repressor binds to multiple short (13-bp) asymmetric repeats (operators and stoperators) located in small intergenic gaps (17, 18) preventing lytic gene transcription in the prophage. The cluster CA *Rhodococcus* phages each have up to 20 copies of repeated sequences related to the 13-bp consensus 5′-TGTCTATTGTCAA, positioned in intergenic spaces and primarily in one orientation with respect to the direction of transcription. This array of features shared between cluster A mycobacteriophages and cluster CA *Rhodococcus* phages suggests they may warrant inclusion in a higher taxonomic level, a "supercluster" (19).

Accession number(s). The 19 *Rhodococcus erythropolis* phage genome sequences are available in GenBank with accession numbers as shown in Table 1.

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, Hatfull GF. 2017. PhagesDB: the actinobacteriophage data-

base. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
3. Salifu SP, Valero-Rello A, Campbell SA, Inglis NF, Scortti M, Foley S, Vázquez-Boland JA. 2013. Genome and proteome analysis of phage E3 infecting the soil-borne actinomycete *Rhodococcus equi*. Environ Microbiol Rep 5:170–178. https://doi.org/10.1111/1758-2229.12028.
4. Summer EJ, Liu M, Gill JJ, Grant M, Chan-Cortes TN, Ferguson L, Janes C, Lange K, Bertoli M, Moore C, Orchard RC, Cohen ND, Young R. 2011.

Genomic and functional analyses of *Rhodococcus equi* phages Reqi-Pepy6, ReqiPoco6, ReqiPine5, and ReqiDocB7. Appl Environ Microbiol 77:669–683. https://doi.org/10.1128/AEM.01952-10.

5. Petrovski S, Dyson ZA, Seviour RJ, Tillett D. 2012. Small but sufficient: the *Rhodococcus* phage RRH1 has the smallest known *Siphoviridae* genome at 14.2 kilobases. J Virol 86:358–363. https://doi.org/10.1128/JVI.05460-11.

6. Petrovski S, Seviour RJ, Tillett D. 2013. Genome sequence and characterization of a *Rhodococcus equi* phage REQ1. Virus Genes 46:588–590. https://doi.org/10.1007/s11262-013-0887-1.

7. Petrovski S, Seviour RJ, Tillett D. 2013. Characterization and whole genome sequences of the *Rhodococcus* bacteriophages RGL3 and RER2. Arch Virol 158:601–609. https://doi.org/10.1007/s00705-012-1530-5.

8. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SC, Findley AM, Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. Söding J. 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics 21:951–960. https://doi.org/10.1093/bioinformatics/bti125.

15. 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.

16. Hatfull GF, Sarkis GJ. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol 7:395–405. https://doi.org/10.1111/j.1365-2958.1993.tb01131.x.

17. Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF. 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. EMBO J 16:5914–5921. https://doi.org/10.1093/emboj/16.19.5914.

18. Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, Alcoser TA, Alexander LM, Alfano MB, Alford ST, Amy NE, Anderson MD, Anderson AG, Ang AA, Ares M, Barber AJ, Barker LP, Barrett JM, Barshop WD, Bauerle CM, Bayles IM, Belfield KL, Best AA, Borjon A, Bowman CA, Boyer CA, Bradley KW, Bradley VA, Broadway LN, Budwal K, Busby KN, Campbell IW, Campbell AM, Carey A, Caruso SM, Chew RD, Cockburn CL, Cohen LB, Corajod JM, Cresawn SG, Davis KR. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One 6:e16329. https://doi.org/10.1371/journal.pone.0016329.

19. Grose JH, Casjens SR. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468–470:421–443. https://doi.org/10.1016/j.virol.2014.08.024.