

Creative Commons Attribution 4.0 International (CC BY 4.0)

<https://creativecommons.org/licenses/by/4.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.



Genomic Sequence and Characteristics of EmiRose, a Bacteriophage Isolated on *Corynebacterium flavescens*

Emily Rose Davis,^a Blesyn Perry-Richardson,^b Aleem Mohamed,^a Ilzat Ali,^a Danielle Heller,^c  Viknesh Sivanathan^c

^aDepartment of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland, USA

^bPublic Safety Program, University of Maryland Global Campus, Adelphi, Maryland, USA

^cDepartment of Science Education, Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

ABSTRACT Bacteriophage EmiRose is a siphovirus infecting *Corynebacterium flavescens*. The EmiRose genome is 37,431 bp long and composed of 47 protein-coding genes. Based on gene content similarity, EmiRose is not closely related to any previously sequenced bacteriophages in the actinobacteriophage database to date, including other corynebacteriophages. EmiRose is classified as a singleton.

Bacteriophages play complex roles in bacterial pathogenesis. In the case of *Corynebacterium diphtheriae*, severe illness is caused by bacterial expression of a prophage-encoded toxin (1). Conversely, bacteriophages are increasingly being considered for use as therapeutics for treating infections caused by diverse bacterial pathogens (2). Given this multifaceted role in medicine, the isolation and genetic characterization of new bacteriophages is important. Here, we report on EmiRose, a corynebacteriophage isolated from topsoil collected from a sidewalk in West Hollywood, CA (global positioning system [GPS] coordinates, 34.076042 N, 118.3639 W), using standard methods (3). Briefly, EmiRose was extracted by washing the soil sample with peptone-yeast extract-calcium (PYCa) liquid medium supplemented with 0.1% dextrose. EmiRose was then enriched in the filtered (0.2- μ m pore size) wash and purified with three successive rounds of plating on *Corynebacterium flavescens* strain ATCC 10340 at 30°C. Top agar overlay of EmiRose resulted in small turbid plaques 1 to 2 mm in diameter after 48 h of incubation at 30°C (Fig. 1A). Negative-stain transmission electron microscopy revealed EmiRose to be a siphovirus with a flexible tail 141 nm long and an isometric capsid 43 nm in diameter (Fig. 1B).

Genomic DNA was isolated from EmiRose using the Promega Wizard DNA cleanup kit, prepared for sequencing using the NEBNext Ultra II FS kit, and sequenced using an Illumina MiSeq instrument, yielding 1,025,624 single-end 150-bp reads constituting 4,110-fold coverage of the genome. The untrimmed reads were assembled as a single contig using Newbler v2.9 and checked for completeness and genomic termini using Consed v2.9 as previously described (4), yielding a genome 37,431 bp long with 9-nucleotide 3' single-stranded cohesive ends (5'-CCCTCCGGT-3') and an average G+C content of 56.4%. Based on its gene content similarity (GCS) of 35% or higher to sequenced bacteriophages present in the actinobacteriophage database, PhagesDB (5), using the PhagesDB GCS tool and previously described criteria (6), EmiRose is not closely related to any bacteriophage, including 19 other corynebacteriophages isolated on *Corynebacterium xerosis* and *Corynebacterium vitæruminis*. EmiRose is thereby classified as a singleton. The genome was annotated using DNA Master (<http://cobamide2.bio.pitt.edu/>), Glimmer v3.02 (7), GeneMark v3.25 (8), BLAST (9), HHpred (10), Aragorn v1.1 (11), and tRNAscan-SE v2.0 (12), all with default parameters. The resulting annotation process revealed a total of 47 protein-coding genes, for which 26 proteins could be assigned putative functions (Fig. 1C). Genes coding for functions in DNA packaging

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2022 Davis 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 Viknesh Sivanathan, sivanathanv@hhmi.org.

The authors declare no conflict of interest.

Received 31 January 2022

Accepted 10 February 2022

Published 2 March 2022

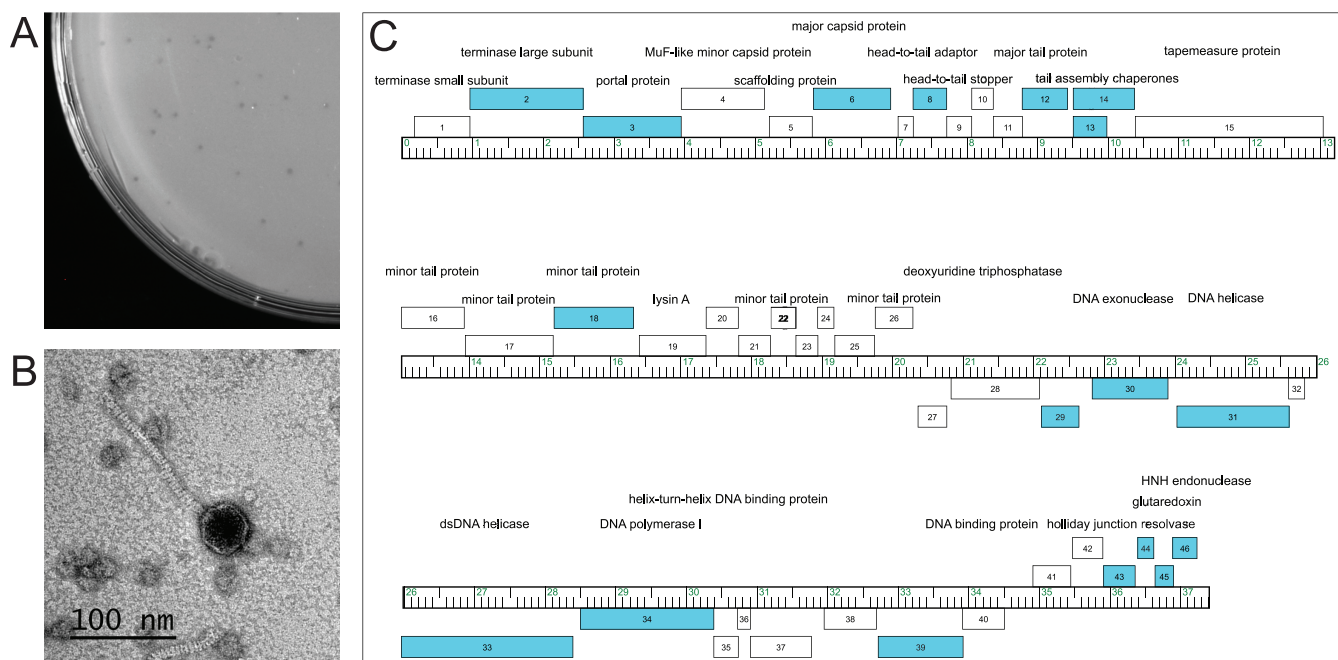


FIG 1 Characteristics of corynebacteriophage EmiRose. (A) EmiRose forms small and turbid plaques 1 to 2 mm in diameter after 48 h of incubation at 30°C. (B) A negative-stain transmission electron micrograph showing the *Siphoviridae* morphology of EmiRose, with a flexible tail 141 to 144 nm long attached to an isometric capsid 43 to 44 nm in diameter, based on measurement of 2 particles. (C) A map of the EmiRose genome, with genes presented as boxes along a genome ruler (in kilobases). The boxes above and below the ruler represent genes that are transcribed rightward and leftward, respectively. The blue and white boxes represent genes coding for proteins with and without homologs shared with other bacteriophages, respectively. The gene number is presented within each box.

(1, 2), virion structure and assembly (3 to 6, 8, 9, 12 to 18, 22, 26), and lysis (18) occupy the left half of the genome and are transcribed rightward. DNA metabolism genes are dispersed throughout the right half of the genome, where the genes are transcribed leftward, although the six rightmost genes are transcribed rightward. EmiRose does not encode identifiable integrase or immunity repressor functions, and it is unlikely to establish lysogeny. Similarly, no identifiable toxin is encoded by EmiRose, and we note that gp4, encoding a MuF-like protein that is toxin associated in some other bacteriophages, lacks the toxin domain. Based on the protein sequence similarity, as previously described (13), only 18 EmiRose proteins have homologs shared with other bacteriophages (<https://phagesdb.org/>). These include 17 proteins with homologs found broadly among actinobacteriophages and 1 (minor tail protein, gp17) shared only with the singleton bacteriophage GMA4, which was isolated on *Gordonia malaquae*.

Data availability. EmiRose is available at GenBank under accession no. [MN586033](#) and Sequence Read Archive (SRA) accession no. [SRX5653831](#).

ACKNOWLEDGMENTS

We thank Daniel Russell and Rebecca Garlena for sequencing and assembling the genome, Deborah Jacobs-Sera for support in annotating the genome, Deborah Jacobs-Sera and Graham Hatfull for feedback on the manuscript, and the SEA-PHAGES program for support.

REFERENCES

- Sharma NC, Efstratiou A, Mokrousov I, Mutreja A, Das B, Ramamurthy T. 2019. Diphtheria. *Nat Rev Dis Primers* 5:81. <https://doi.org/10.1038/s41572-019-0131-y>.
- Hatfull GF, Dedrick RM, Schooley RT. 2022. Phage therapy for antibiotic-resistant bacterial infections. *Annu Rev Med* 73:197–211. <https://doi.org/10.1146/annurev-med-080219-122208>.
- Poxleitner M, Pope W, Jacobs-Sera D, Sivanathan V, Hatfull GF. 2018. HHMI SEA-PHAGES phage discovery guide. <https://seaphagesphagediscoveryguide.helpdocsonline.com/home>.
- Russell DA, Hatfull GF. 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.
- Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. *Bioinformatics* 33:784–786. <https://doi.org/10.1093/bioinformatics/btw711>.
- 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>.
7. 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>.
 8. Borodovsky M, McIninch J. 1993. Recognition of genes in DNA sequence with ambiguities. *Biosystems* 30:161–171. [https://doi.org/10.1016/0303-2647\(93\)90068-N](https://doi.org/10.1016/0303-2647(93)90068-N).
 9. 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](https://doi.org/10.1016/S0022-2836(05)80360-2).
 10. 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>.
 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, 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>.
 13. 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>.