

The complete genome sequence of the *Microbacterium* bacteriophage Erla

Running title: Genome of the bacteriophage Erla

Mark Milhaven^a, Erin Hastings^b, Danielle Brister^a, Leo Cevallos^a, Sriya Chilukuri^a, Arshia Diyya^a, Aman Garg^a, Rosaura Hernandez^a, Daniel Kelly^a, Karina Lazo^a, Jennifer Le^a, Garrett Maag^a, Palak D. Marfatia^a, Liam McCarthy^{b,c}, Rithik Mehta^a, Aram Nejad^{a,d}, Jade Porche^a, Alexander Queiroz^a, Daniel Sackett^a, Pablo Santos Molina^a, Taylor Slade^a, Minerva So^a, Karan Thakur^b, Angelica Urquidez Negrete^a, Sage Wackett^a, Sarah Weiss^a, Adam D. Rudner^{b,e}, John P. McCutcheon^{a,d,†}, and Susanne P. Pfeifer^{a,d,f,†}

^a School of Life Sciences, Arizona State University, Tempe, Arizona, United States

^b Translational and Molecular Medicine Program, University of Ottawa, Ottawa, Ontario, Canada

^c Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, Canada

^d Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, United States

^e Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, Canada

^f Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, United States

[†] These authors jointly supervised the project

Corresponding author: Susanne P. Pfeifer, susanne.pfeifer@asu.edu

We characterized the complete genome of the *Siphoviridae* Erla, an obligatory lytic subcluster EA1 bacteriophage infecting *Microbacterium foliorum* NRRL B-24224, with a capsid width of 65nm and a tail length of 112nm. The 41.5kb genome, encompassing 62 predicted protein-coding genes, is highly similar (99.52% identity) to that of bacteriophage Calix.

The discovery of antibiotic agents in 1928 by Alexander Fleming has completely changed the way clinicians treat bacterial infections. However, with new therapeutic challenges resulting from antibiotic resistance, scientists have now turned to bacteriophages (*i.e.*, viruses that infect bacteria) as promising clinical alternatives to antibiotics [1]. Aiding this endeavor is the vast – though largely uncharacterized – diversity of bacteriophages across the globe (current estimates range in the order of 10³¹[2]).

Here, we report the complete genome sequence of the bacteriophage Erla, obtained from a garden in Ottawa, Ontario, Canada (45.423407N, 75.682795W). This bacteriophage, sequenced as part of Howard Hughes Medical Institute (HHMI) Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Sciences (SEA-PHAGES) undergraduate research program, was isolated at 16°C, purified, and

amplified in *Microbacterium foliorum* NRRL B-24224 (a gram-positive strain isolated from the phyllosphere of grasses in Germany, commonly used as a host for bacteriophage discovery [3]), following the standard procedures outlined in the SEA-PHAGES Discovery Guide (<https://seaphagesphagediscoveryguide.helpdocsonline.com/home>). Erla forms medium-sized plaques with a faint bullseye in the middle (Fig. 1a), and exhibits a *Siphoviridae* morphology with an icosahedral capsid (diameter: 65nm) enclosing the double-stranded DNA, attached to a flexible, non-contractile tail (length: 112nm; Fig. 1b).

In order to sequence Erla's genome, DNA was extracted from the phage lysate using phenol/chloroform, followed by ethanol precipitation. Next, a sequencing library was prepared using the NEB Ultra II FS kit and sequenced on an Illumina MiSeq (150bp single-end reads) to >3,300X coverage. Following [4], quality-trimmed reads were *de novo* assembled using Newbler v.2.9 [5], resulting in a single linear contig of size 41,538bp with a GC-content of 63.4% – similar to the 68.7% GC-content of its host, *Microbacterium foliorum* NRRL B-24224. The assembly was checked for completeness, accuracy, and genome termini using Consed v.29.0 [6].

Following the HHMI SEA-PHAGES Phage Genomics Guide (<https://seaphagesbioinformatics.helpdocsonline.com/home>), Erla's genome was

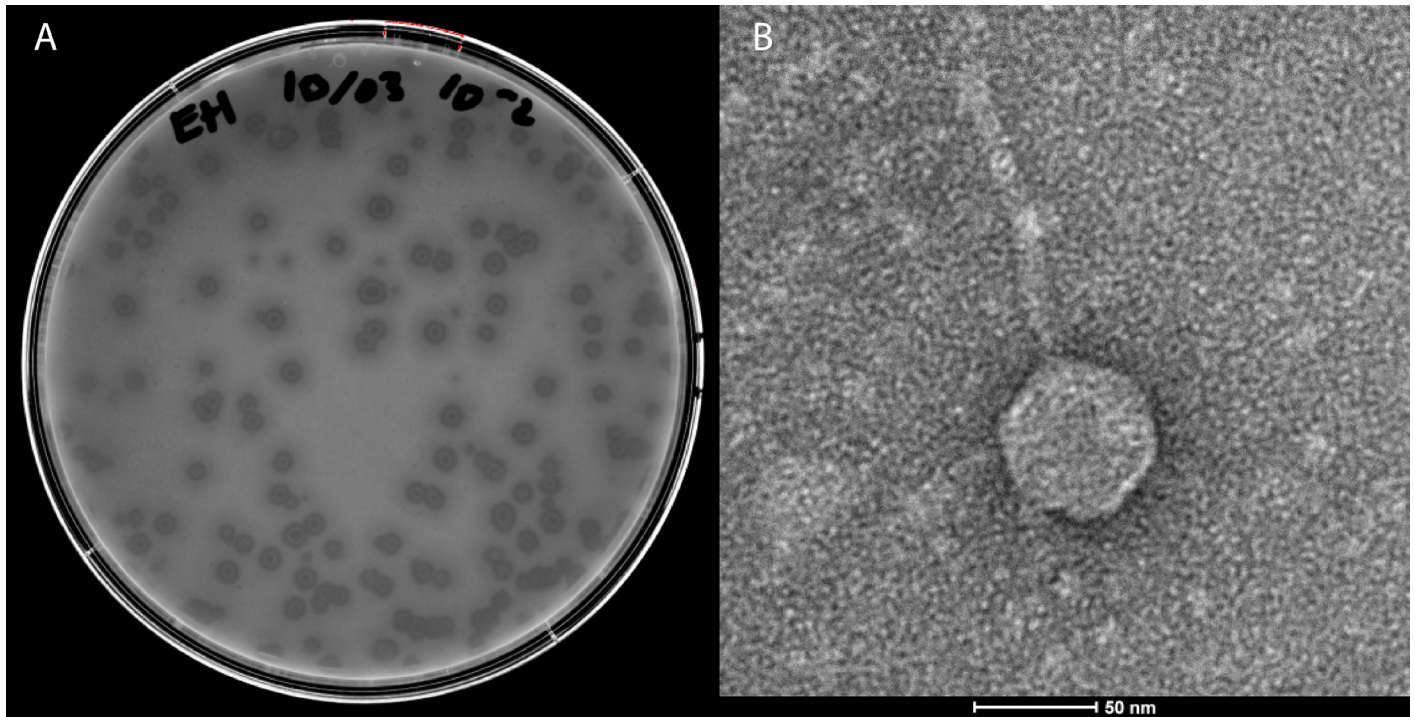


FIG. 1 Characterization of the *Microbacterium* bacteriophage Erla. A) Erla forms medium-sized plaques with a faint bullseye in the middle. B) Transmission electron microscopy image showing Erla's *Siphoviridae* morphology with an icosahedral capsid (diameter: 65nm) attached to a 112nm-long tail.

annotated using DNAMaster v.5.23.5 (<http://cobamide2.bio.pitt.edu>). Sixty-two putative genes were identified using Glimmer v.3.02 [7], GeneMark v.3.25 and v.4.28 [8], and Starterator v.381 (<https://seaphages.org/software>), corresponding to a gene density of 1.49 genes/kb. ARAGORN v.1.1 (included in DNAMaster) and v.1.2.38 [9] as well as tRNAscan-SE v.2.0 [10] were used to search for tRNAs and tmRNA, but none were found. Functional assignments were made using BLASTp v.2.10.1 [11] and HHpred [12], leading to a putative function for 26 out of the 62 genes. TMHMM v.2.0 [13] and SOSUI v1.11 [14] were used to gather further information on proteins of no known function, leading to the identification of two additional transmembrane proteins.

Phamerator v.381 [15] was used to determine synteny among Erla and other bacteriophages previously sequenced as part of the SEA-PHAGES program, which supported the above functional assignments. Multiple sequence alignments using Kalign v.1.04 [16] and BLASTn v.2.10.1 [12] indicated that Erla is a *Microbacterium* subcluster EA1 bacteriophage, most closely related to Calix (percent identity: 99.52%; GenBank accession number: MN234163.1), Gelo (percent identity: 99.51%; GenBank accession number: MG962367.1) and Etta (percent identity: 99.51%; GenBank accession number: MK977697.1).

DATA AVAILABILITY

Whole genome sequencing data are available through NCBI's Sequence Read Archive (BioProject accession number XXX) and the annotated genome assembly is available through GenBank (accession number XXX).

ACKNOWLEDGEMENTS

This study was supported by the Howard Hughes Medical Institute SEA-PHAGES program and Arizona State University's School of Life Sciences. Computations were partially performed at Arizona State University's High Performance Computing facility. We are grateful to Suhail Ghafoor for IT support, Billy Biederman, Graham Hatfull, Deborah Jacobs-Sera, Welkin Pope, Daniel Russell, and Vic Sivanathan for library preparation, sequencing, and assembly as well as providing faculty training for the SEA-PHAGES program and guidance with our genome annotation.

REFERENCES

1. Romero-Calle D, Guimarães Benevides R, Góes-Neto A, Billington C. 2019. Bacteriophages as alternatives to antibiotics in clinical care. *Antibiotics (Basel)*, 8(3): 138.
2. Rohwer F. 2003. Global phage diversity. *Cell*, 113(2): 141.
3. Russell DA, Garlena RA, Hatfull GF. 2019. Complete genome sequence of *Microbacterium foliorum* NRRL B-24224, a host for bacteriophage discovery. *Microbiol. Resour. Announc.*, 8(5): e01467-18.

4. Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes. *Methods Mol. Biol.*, 1681: 109–125.
5. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057): 376–380.
6. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8(3): 195–202.
7. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.*, 27(23): 4636–4641.
8. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.*, 26(4): 1107–1115.
9. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, 32(1): 11–16.
10. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5): 955–964.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215(3): 403–410.
12. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33: W244–W248.
13. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* 305(3): 567–580.
14. Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14(4): 378–379.
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.
16. Lassmann T, Sonnhammer EL. 2005. Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*, 6: 298.