

Complete Genome Sequences of 3 Newly Isolated C1 Mycobacteriophages: Erdmann, Yucca,
and Daffodil

Madeline D. Gibson,¹ Brooke E. Koebele,¹ R. L. Shafer,¹ Emily A. Erdmann,¹ Niyant S. Vora,¹
Addison D. Ely,¹ Ellen R. Stumph,¹ Julia M. Lennon,¹ Daniel J. Walski,¹ Megan E. Dolan,¹
Alexandria S. Paradis,¹ Samridh Gupta,¹ Seth H. Borrowman,¹ Morgan A. Braun,¹ Richard M.
Alvey,¹ David W. Bollivar,¹ Daniel A. Russell,² and Graham F. Hatfull²

¹Illinois Wesleyan University, Department of Biology, Bloomington, IL 61701

²Pittsburgh Bacteriophage Institute, Department of Biological Sciences, Pittsburgh, PA 15260

Abstract

Three bacteriophages (Daffodil, Erdmann, and Yucca) that infect *Mycobacterium smegmatis* mc²155 were isolated from soil in Bloomington, Illinois. All three bacteriophages belong to the C1 cluster of mycobacteriophages and to group Myoviridae. Currently, 97 C1 mycobacteriophages have been sequenced (<http://phagesdb.org/>); thus, these novel phages comprise 3.09% of all known C1 phages.

Genome Announcement

Bacteriophages, found ubiquitously in large numbers, provide insight into evolutionary science and show potential for use in the medical field (1). Bacteriophages are classified by which genus of bacteria they infect, and are then grouped into progressively smaller, more specific clusters. A cluster is defined as a group of phages sharing nucleotide similarity over at least 50% of their genomes (1).

Yucca, Erdmann, and Daffodil were isolated from soil in Bloomington, Illinois. Yucca was isolated using the direct plating method, and Daffodil and Erdmann were isolated by enrichment (2). Plaques of Yucca, Erdmann, and Daffodil were all relatively slow-appearing, usually presenting within 48 hours. Plaques were not completely clear and were relatively small, ranging in size from 0.5 to 1.5 mm. Following isolation of DNA from all three phages, they were sent to the University of Pittsburgh for sequencing. Daffodil, Erdmann, and Yucca were sequenced using Illumina MiSeq and assembled with 140 bp single end reads using Newbler. Daffodil, Erdmann, and Yucca have double stranded, circular genomes, consisting of 155,034 base pairs, 155,565 base pairs, and 155,582 base pairs, respectively. The fold coverages were 1206X, 319X, and 110X, for Daffodil, Erdmann, and Yucca, respectively. All three phages have a GC content of 64.7%. Upon receiving the sequenced genomes, the computer programs DNAMaster (available at: <http://cobamide2.bio.pitt.edu/>), Aragorn (3), GeneMark (4), Glimmer (5), Phamerator (6), HHPred (7), and NCBI Blastp (8) were used to analyze and annotate the genomes, as described in the second portion of the HHMI SEA PHAGES program manual (2). Yucca, Erdmann, and Daffodil all belong to cluster C1 and have 233, 234, and 238 predicted genes, respectively. Of the predicted genes, functions were assigned to 48 genes in Daffodil, 59 genes in Erdmann, and 60 genes in Yucca. All three phages were found to possess 35 tRNA genes and 1 tmRNA gene.

In 2011, Illinois Wesleyan SEA-PHAGES students isolated two C1 phages, Shrimp and Gizmo. Both of these phages proved useful in the annotation processes of Daffodil, Erdmann, and Yucca, as their genomes are nearly identical. Shrimp shares 99% identity with Yucca, Erdmann, and Daffodil, spanning 97%, 99%, and 98 % of their genomes, respectively. Gizmo shares 99 % identity with Yucca and Erdmann spanning 98% of their genomes and 99% identity with Daffodil spanning 99% of the genome. The proximity of collections sites could explain this great homology; Shrimp, Gizmo, Daffodil, Erdmann, and Yucca were all isolated within 1.5 km (<1 mile) of each other. All three phages also share great homology with phages Ava3, found in Grand Rapids, Michigan, Pleione, found in St. Louis, Missouri, and Willis, found in Pittsburgh, Pennsylvania (<http://phagesdb.org/>).

Yucca, Erdmann, and Daffodil all share 99% identity spanning 98% of their genomes. Despite the great similarity of the genomes of Yucca, Erdmann, and Daffodil, certain differences are particularly intriguing. For example, gene 263 of Yucca shows no homology with any other genes previously discovered. Yucca is also the only one of the three novel phages to have an endonuclease VII gene. While Shrimp, Gizmo, Yucca, and Erdmann all have an identical HNH homing endonuclease gene, Daffodil does not.

Accession numbers: Erdmann, Yucca, and Daffodil genome sequences are available from GenBank under the accession numbers: KX721256, KX721255, and MF919499, respectively.

Acknowledgements:

IWU, HHMI SEA PHAGES, University of Pittsburgh

References

1. **Hatfull GF**. 2012. The secret lives of mycobacteriophages. *Adv Virus Res* 82:179–288. doi:10.1016/B978-0-12-394621-8.00015-7 [[PubMed](#)] [[Cross Ref](#)]
2. **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. doi:10.1128/mBio.01051-13.
3. **Laslett D, Canback B**. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res*. 2004;32:11–6. [[PMC free article](#)] [[PubMed](#)]
4. **Besemer J, Lomsadze A, Borodovsky M**. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618. 10.1093/nar/29.12.2607. [Abstract/FREE Full Text](#)
5. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL**. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*. 1999;27:4636–41. [[PMC free article](#)] [[PubMed](#)]
6. **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. doi:10.1186/1471-2105-12-395 [[PMC free article](#)] [[PubMed](#)] [[Cross Ref](#)]
7. **Söding J, Biegert A, Lupas AN**. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248. 10.1093/nar/gki408. [Abstract/FREE Full Text](#)
8. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. 10.1016/S0022-2836(05)80360-2. [CrossRefMedlineGoogle Scholar](#)