Genome Analysis of *Microbacterium foliorum* Bacteriophage Annapurna

Ashtyn Stock[1*], Sarah Pritchard[1*], Shelby Stone[1*], Kinley Shelton[1*], Emme Clowdus[1*], Emily Doughtery[1*], Sallie Hicks[1*], Annabelle McDaniel[1*], Kelsey Towle[1*], Hannah Joseph[1*], Kati Pourfarzib[1*], Michael Baker[1*], William Myers[1*], Jadin Allen[1*], Sydney Kruger[1*], Kevin Perez[1*], Jessie Holsombeck[1*], Lori West[1*], Dana Perry[1*], Joseph Daft[1§*]

[1]Natural Sciences, Lee University, Cleveland, Tennessee, United States

[§]To whom correspondence should be addressed: jdaft@leeuniversity.edu

[*]These authors contributed equally.

Abstract

Bacteriophage Annapurna is a siphovirus discovered within a soil sample collected in North Georgia. Annapurna was isolated and amplified using the host *Microbacterium foliorum* prior to genome sequencing. Annapurna contains 84 predicted protein-coding genes encoded across a genome 56,247 base pairs in length. Based on gene content, Annapurna is assigned to actinobacteriophage cluster EF. Like other phages in the EF cluster, Annapurna appears to lack both integrase and repressor genes and is therefore predicted to be a virulent phage.
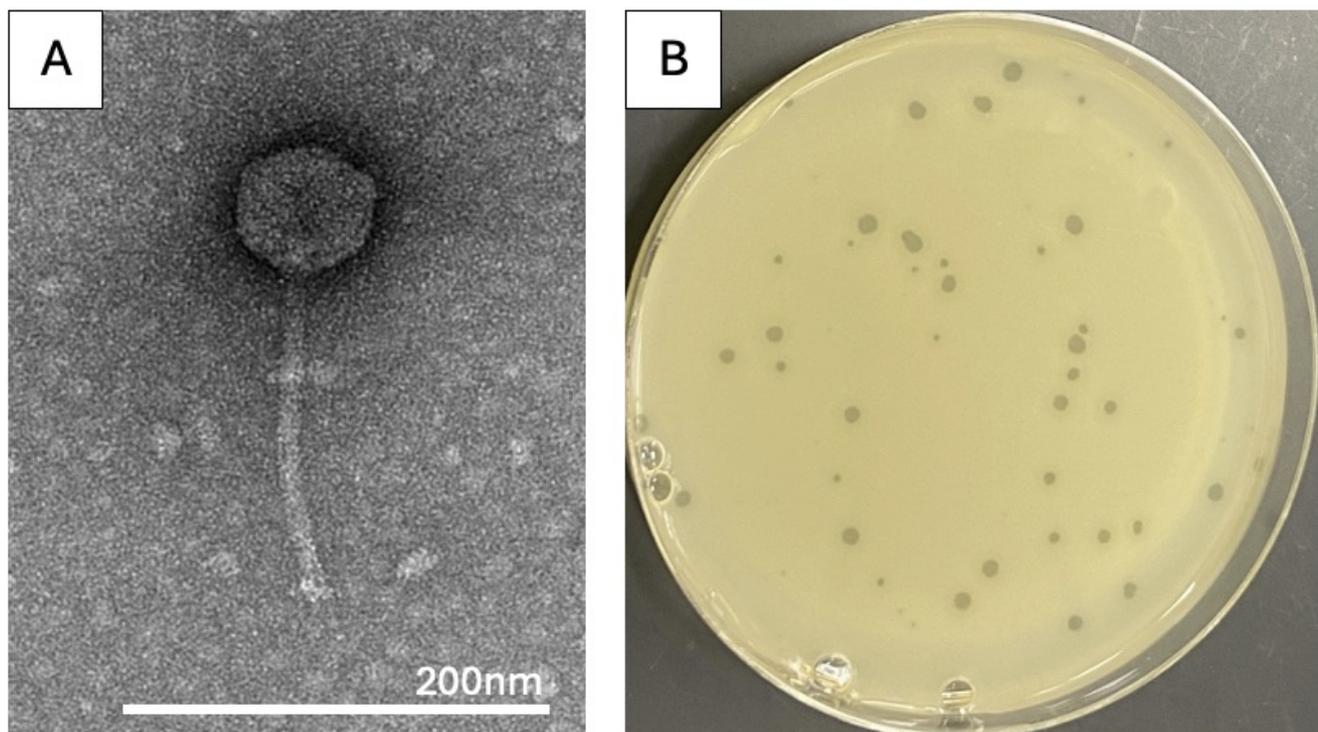
Figure 1. Transmission electron micrograph of an Annapurna virion and plaque morphology:

A) Transmission electron micrograph of Annapurna showing a capsid diamter of 67-72nm and a 145-153nm tail (n=4), imaged at the University of Maryland Baltimore County. B) Plaque morphology of Annapurna plated with *Microbacterium foliorum.*

Description

Due to the increase in antibiotic-resistant bacterial infections, there is a need for alternative treatment options. A promising avenue for treatment is the use of bacteriophages (Hatfull, 2022). In this study, *Microbacterium foliorum* NRRL B-24224, a common soil bacterium, was used as a host to isolate and purify a novel bacteriophage. The bacteriophage Annapurna was isolated from a dry soil sample collected from a cow pasture in Catoosa County, Georgia, USA (global positioning system [GPS] 34.56015N, 85.02109W) as previously described (Zorawik et al, 2024). Briefly, the soil sample was washed with peptone-yeast extract-calcium (PYCa) liquid media and the wash then filtered through a 0.22-µm filter. The filtrate was then plated in top agar with *M. foliorum* and plates incubated at 30°C for 48 hours. A clear circular plaque, with a mean diameter of 1.90mm ± 0.78mm (n=20), was picked and designated Annapurna (Figure 1). Multiple

rounds of plating were performed to purify Annapurna, after which a lysate was prepared and used for imaging and DNA extraction. Negative-stain (1% uranyl acetate) transmission electron microscopy (TEM) showed Annapurna to have a siphovirus morphology. (Figure 1).

Annapurna's DNA was extracted using the Promega Wizard DNA cleanup kit, prepared for sequencing using the NEB Ultra II FS kit, and sequenced at the University of Pittsburgh using Illumina NextSeq 1000 (XLEAP-P1 kit). The resulting 100-base reads were trimmed with cutadapt 4.7 (using the option: –nextseq-trim 30) (Martin, 2011) and filtered with skewer 0.2.2 (using the options: -q 20 -Q 30 -n -l 50) prior to assembly with Newbler v2.9 (Margulies et al., 2005) and assembly with 3,761-fold coverage checked for completeness with Consed v29 (Gordon et al., 1998). Annapurna's genome is 56,247 base pairs in length and has 63.5% GC content with circularly permuted ends. Based on gene content similarity equal to or greater than 35% to phages in the Actinobacteriophage database, PhagesDB, Annapurna is assigned to the EF cluster (Pope et al., 2017; Russell and Hatfull, 2017).

The genome of Annapurna was annotated utilizing Glimmer v3.02 and GeneMark v4.28 (Besemer and Borodovsky, 2005; Delcher, et al., 2007) to identify potential gene coding regions. These predictions were reviewed and refined in DNA Master v5.23.6 before functional predictions using amino acid sequences were assigned using HHpred MPI Bioinformatics Toolkit v3.3, using the PDB_mmCIF70, Pfam-v.37., NCBI Conserved Domains v3.21, and Uniprot SwissProt Viral70 databases (Söding et al., 2005; Pope and Jacobs-Sera, 2018). BLAST, using the Actinobacteriophage and NCBI non-redundant database, was used to confirm possible functions based on similarity with known sequences (Altschul et al., 1990). PhagesDB and Phamerator, with Actino_draft database v578, were used to compare Annapurna's genome with other phages in the EF cluster (Cresawn, 2011; Russell and Hatfull, 2016). DeepTMHMM v1.0.44 was used to identify transmembrane proteins while Aragon v1.2.41 and tRNAscan-SE v2.0.6 were used to identify tRNA sequences (Lowe and Eddy, 1997; Laslett and Canback, 2004). Default settings were used for all software during genome annotation. The annotation process revealed 84 predicted protein-coding genes, no tRNAs, and no tmRNAs. Of the protein-coding genes, 26 were assigned putative functions while 7 were identified as potential membrane proteins. The remaining genes code for proteins with no known function.

As with all 42 other annotated EF phages to-date, all predicted genes in Annapurna are transcribed unidirectionally. Additionally, there are two genes encoding for DnaE-like DNA polymerase alpha subunits and no integrase and repressor genes were identified, suggesting EF phages are unable to establish lysogeny (Jacobs-Sera, et al., 2020).

Nucleotide sequence accession numbers

Annapurna is available at GenBank with Accession No. PV915857 and Sequence Read Archive (SRA) No. SRX29714288.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410. DOI: 10.1016/S0022-2836(05)80360-2

Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Research 33: W451-W454. DOI: 10.1093/nar/gki487

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: 10.1186/1471-2105-12-395. DOI: 10.1186/1471-2105-12-395

Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23: 673-679. DOI: 10.1093/bioinformatics/btm009

Gordon D, Abajian C, Green P. 1998. *Consed:* A Graphical Tool for Sequence Finishing. Genome Research 8: 195-202. DOI: 10.1101/gr.8.3.195

Hatfull GF. 2022. Mycobacteriophages: From Petri dish to patient. PLOS Pathogens 18: e1010602. DOI: 10.1371/journal.ppat.1010602

Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, Bhalla SS, et al., Hatfull. 2020. Genomic diversity of bacteriophages infecting Microbacterium spp. PLOS ONE 15: e0234636. DOI: 10.1371/journal.pone.0234636

Laslett D. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Research 32: 11-16. DOI: 10.1093/nar/gkh152

Lowe TM, Eddy SR. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Research 25: 955-964. DOI: 10.1093/nar/25.5.955

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al., Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-380. DOI: 10.1038/nature03959

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10. DOI: 10.14806/ej.17.1.200

Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, et al., Hatfull. 2017. Bacteriophages of *Gordonia* spp. Display a Spectrum of Diversity and Genetic Relationships. mBio 8: 10.1128/mbio.01069-17. DOI: 10.1128/mbio.01069-17

Pope WH, Jacobs-Sera D. 2017. Annotation of Bacteriophage Genome Sequences Using DNA Master: An Overview. Methods in Molecular Biology,Bacteriophages : 217-229. DOI: 10.1007/978-1-4939-7343-9_16

Russell DA, Hatfull GF. 2016. PhagesDB: the actinobacteriophage database. Bioinformatics 33: 784-786. DOI: 10.1093/bioinformatics/btw711

Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33: W244-W248. DOI: 10.1093/nar/gki408

Zorawik M, Jacobs-Sera D, Freise AC, SEA-PHAGES, Reddi K. 2024. Isolation of Bacteriophages on Actinobacteria Hosts. Methods in Molecular Biology,Phage Engineering and Analysis : 273-298. DOI: 10.1007/978-1-0716-3798-2_17

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: Ashtyn Stock: investigation, writing - review editing. Sarah Pritchard: investigation, writing - original draft. Shelby Stone: investigation, writing - original draft. Kinley Shelton: investigation, writing - original draft. Emme Clowdus: writing - original draft. Emily Doughtery: investigation, writing - original draft. Sallie Hicks: investigation, writing - original draft. Annabelle McDaniel: investigation, writing - original draft. Kelsey Towle: investigation, writing - original draft. Hannah Joseph: investigation, writing - original draft. Kati Pourfarzib: investigation, writing - original draft. Michael Baker: investigation, writing - original draft. William Myers: investigation, writing - original draft. Jadin Allen: investigation, writing - original draft. Sydney Kruger: investigation, writing - original draft. Kevin Perez: investigation, writing - original draft. Jessie Holsombeck: validation, writing - review editing. Lori West: validation, writing - review editing. Dana Perry: investigation, visualization. Joseph Daft: validation, writing - review editing.

Reviewed By: Dustin Edwards, Maria Diane Gainey Dr.

Citation: Stock A, Pritchard S, Stone S, Shelton K, Clowdus E, Doughtery E, et al., Daft J. 2026. Genome Analysis of *Microbacterium foliorum* Bacteriophage Annapurna. microPublication Biology. 10.17912/micropub.biology.001947