Genome Sequences of Two F1 Subcluster Bacteriophages, Burwell21 and Nivrat, Isolated Using the Bacterial Host Mycobacterium smegmatis

Joanna Katsanos,a Chioma Ngene,b Jennifer Cook-Easterwooda

aDepartment of Biology, Queens University of Charlotte, Charlotte, North Carolina, USA
bDepartment of Chemistry, Queens University of Charlotte, Charlotte, North Carolina, USA

ABSTRACT The mycobacteriophages Burwell21 and Nivrat are two F1 cluster bacteriophages isolated from different soil samples in Charlotte, NC. Burwell21 has a 58,098-base-pair double-stranded DNA genome, with 99 protein-coding genes predicted, whereas Nivrat has a 58,009-base-pair genome, with 102 protein-coding genes predicted.

We report the genome sequences of two F1 subcluster mycobacteriophages, Burwell21 and Nivrat, that were isolated from two different soil samples (35.288°N, 80.832°W and 35.189°N, 80.834°W, respectively). Both soil samples were collected from flower beds (rich, dark soil) on the campus of Queens University of Charlotte in North Carolina as part of the Howard Hughes Medical Institute (HHMI) Science Education Alliance-Phage Hunters Advancing Genomic and Evolutionary Science (SEA-PHAGES) program (1). The isolation, characterization, genome sequencing, and annotation of Burwell21 and Nivrat contribute to a greater understanding of the diversity and abundance of mycobacteriophage species (1).

We used the protocols provided in the HHMI SEA-PHAGES discovery guide (https://seaphagesphagediscoveryguide.helpdocsonline.com/home) to isolate the mycobacteriophages by enrichment using the bacterial host Mycobacterium smegmatis mc² 155 (provided by the Bacteriophage Institute at University of Pittsburgh) grown in 7H9 complete liquid medium at 37°C for 5 days. Isolation was followed by multiple rounds of plaque assays to purify the phage and the generation of high-volume lysates to amplify the phage. Subsequent negative-staining transmission electron microscopy revealed that Burwell21 and Nivrat have siphoviral morphology. Burwell21 and Nivrat both have a 200-nm tail; however, the diameters of their capsids were 67 nm and 57 nm, respectively.

Genomic DNA was isolated from the bacteriophage using the Promega Wizard DNA cleanup kit and prepared for sequencing with a NEB Ultra II DNA kit. Genomes were sequenced at the Pittsburgh Bacteriophage Institute with an Illumina MiSeq instrument, yielding at least 720,000 reads of 150-base read length per genome (Table 1). The raw reads were assembled using Newbler 2.7 (2) with default settings in each case, yielding a single contig for each phage. Consed v29 (3) was used to check for completeness, accuracy, and genomic termini, as described by Russell (4). Burwell21 had a genome size of 58,098 base pairs, with a fold coverage of 1,753×, a G+C content of 61.5%, and 10-base single-stranded 3′ extensions (5′-CGGTAGG CGC-3′). Nivrat had a genome length of 52,152 base pairs, with a fold coverage of 1,768×, and also had a G+C content of 61.5% with 10-base single-stranded 3′ extensions (5′-CGGAAGGCGC-3′).

The protein-coding regions of the phage genomes were predicted using GLIMMER v3.02 (5) and GeneMark v2.5p (6) embedded within DNA Master v5.02 (http://
TABLE 1 Phage accession numbers and genome assembly results

| Phage name | GenBank accession no. | SRA accession no. | No. of reads | Avg coverage (X) | Cluster | Genome length (bp) | G+C content (%) | No. of genes |
|------------|-----------------------|-------------------|-------------|------------------|---------|-------------------|----------------|-------------|
| Burwell21  | MH651169              | SRX5572879        | 720,592     | 1,753            | F1      | 58,098            | 61.5           | 99          |
| Nivrat     | MH651183              | SRX5572868        | 724,515     | 1,768            | F1      | 52,152            | 61.5           | 102         |

Clustal Omega alignment (11) using default settings between Burwell21 and Nivrat showed a 92.8% average nucleotide sequence identity (ANI) between them. Subsequent Clustal Omega multiple alignments using default settings indicated a 95.5% ANI between Burwell21 and Cabrinians (GenBank accession number KT895281), whereas Nivrat had a 90.4% ANI with Cabrinians. Burwell21 and Nivrat also have 85.3% and 86.1% ANI, respectively, with Saal (GenBank accession number KJ025956). Cabrinians and Saal are both members of the F1 subcluster; hence, Burwell21 and Nivrat are also members of the F1 mycobacteriophage subcluster.

**Data availability.** The GenBank and SRA accession numbers for Burwell21 and Nivrat are listed in Table 1.

**ACKNOWLEDGMENTS**

This work was supported by the Howard Hughes Medical Institute SEA-PHAGES program.

Soil samples were collected and phage isolation was performed in the fall of 2017 by students enrolled in the BIO 201 courses at Queens University of Charlotte. Electron microscopy was performed at the Electron Microscopy Laboratory at Winthrop University. Phage genomes were sequenced at the Pittsburgh Bacteriophage Institute at the University of Pittsburgh. Preliminary genome annotation was performed in the spring of 2018 by the students enrolled in the BIO 303 courses at Queens University of Charlotte.

**REFERENCES**

1. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, Dejong RJ, Donenhy JJ, Denver DR, Dunbar D, Elgin SCR, Findley AM, Gissendanner CR, Golebiowska 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, Jacobus-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.

2. Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next generation sequencing data. Genomics 95:315–327. https://doi.org/10.1016/j.ygeno.2010.03.001.

3. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics 29:2936–2937. https://doi.org/10.1093/bioinformatics/btt515.

4. Russell DA. 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.

5. 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.

6. 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.

7. 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.

8. 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.

9. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupsa AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

10. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.

11. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:239. https://doi.org/10.1038/msb.2011.75.