



Genome Sequence of Mycobacteriophage Bassalto

 Nazir Barekzi,^a Meagan N. Wilkins,^a Aumon L. Williams,^a Afiya J. Moore,^a Zachary R. Duckett,^a Danielle M. Tindall,^a Donnetta R. Eaddy,^a Mary B. Johnson,^a Malcolm Bass,^a  Catherine M. Mageeney^b

^aDepartment of Biology, Norfolk State University, Norfolk, Virginia, USA

^bBiotechnology and Bioengineering, Sandia National Laboratories, Livermore, California, USA

ABSTRACT Bassalto is a newly isolated phage of *Mycobacterium smegmatis* mc²155 from the campus grounds of Norfolk State University in Norfolk, VA. Bassalto belongs to the cluster B and subcluster B3 mycobacteriophages, based on the nucleotide composition and comparison to known mycobacteriophages.

A significant number of *Mycobacterium* phage genomes have been annotated to understand their genomic structure, diversity, and evolution. Further study of phages contributes to the advancement of our general knowledge of genetics, biodiversity, and microbial ecology, which can be applied to environmental science and investigation of phage therapy to combat microbial pathogens such as *Mycobacterium* species (1).

Mycobacteriophage Bassalto was isolated from soil collected from the grounds of Norfolk State University in Norfolk, VA (36.8490817 N, 76.2593061 W), using the enriched isolation protocol in the Phage Discovery Guide (<https://seaphagesphagediscoveryguide.helpdocsonline.com/home>; accessed September 2017), with the host bacterium, *Mycobacterium smegmatis* mc²155. Phages were amplified using the web plate method and flooded with phage buffer followed by filtration (0.2- μ m polytetrafluoroethylene [PTFE] VWR filter) of the supernatant to produce a high-titer lysate (HTL) containing greater than 10⁹ PFU/mL. Subsequently, phage DNA was extracted using the Promega Wizard DNA cleanup kit.

The genome of Bassalto was sequenced to at least 50 \times coverage using the Illumina MiSeq (2 \times 250-bp) platform. Library preparation included tagmentation using the Illumina DNA prep tagmentation kit, followed by optimization with the unique dual index adapters from IDT for the Illumina Nextera DNA unique dual indexes (set C) kit. Sequencing yielded 242,627 total reads. The reads were trimmed using adapter trimming, and quality metrics were obtained using the Illumina FASTQ toolkit v1.1.0. The reads were assembled using Celera WGS v8.1 into a single 68,984-bp scaffold. Upon close inspection of the resulting scaffold, the assembled genome had three unresolved or unknown nucleotide blocks (N-blocks). To resolve the N-blocks, samples of the HTL were used with specific primers designed to flank each N-block by \sim 100 bp and PCR amplified (primers listed in Table 1). Subsequently, the DNA was run on an agarose gel; the fragment was extracted and sent out for sequencing by Genewiz, Inc.

The complete resolved Bassalto genome was 69,113 bp with a G+C content of 67.5%. The coding DNA sequence predictions for Bassalto were determined using Web-based GeneMarkS (2) and command-line multiPhATE v0.5 (3) with the Glimmer v3.02 (4), Prodigal v2.6.3 (5), and PHANOTATE v0.13.0 (6) gene callers turned on. Open reading frames (ORFs) predicted using these gene callers were functionally annotated by conducting Web-based blastp (7) and HHpred (8) searches (using the databases PDB_mmCIF70 and Pfam-A_v35). The genome was run through Web-based tRNAscan-SE (9), and no tRNA genes were predicted. The genome had 99% nucleotide match to cluster B and subcluster B3 mycobacteriophages from the Actinobacteriophage Database, with similar sequence identity to

Editor John J. Dennehy, Queens College
Department of Biology

Copyright © 2023 Barekzi 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 Catherine M. Mageeney, cmmagee@sandia.gov.

The authors declare no conflict of interest.

Received 7 February 2023

Accepted 4 March 2023

Published 21 March 2023

TABLE 1 Primers used to resolve N-blocks

Primer name	Primer sequence (5'–3')	Bassalto genome coordinates
BN1-up	TCGCCGAACGTCAGGCCGAGGAGGG	24741–24765
BN1-down	CCACCAGGGCGTGCGAGTCGTG	25238–25216
B-N2-up	TCATGTCGCTGCCCGTAGG	37770–37788
B-N2-down	CCGAGTTCTTCATGACC	38153–38137
B-N3-up	CTCGTCAGCCGTGGCATCAC	38422–38441
B-N3-down	GTCGCGGCTGAGCACGTG	38546–38529

Phaedrus, OrangeOswald, and Phylar (10), placing it in the class Caudoviricetes and genus *Pegunavirus*.

Bassalto contains 103 ORFs. There was a lack of defined ends, which is consistent among cluster B mycobacteriophages (11). In addition, the latter portion of Bassalto's genome contains coding DNA sequences (CDSs) that encode proteins which function in DNA binding, host identification, and hydrolase inhibition. Notably, Bassalto contains two DNA polymerase genes (ORF95 and ORF98). There is also a notable region at the 3' end of the genome (position 64899 to 65593) that has no nucleotide similarity to any sequence at NCBI and encodes a predicted HNH endonuclease.

Data availability. The Bassalto genome sequence is available at GenBank under accession number [OP777409](https://www.ncbi.nlm.nih.gov/nuclseq/OP777409). The raw reads were deposited at the NCBI Sequence Read Archive (SRA) under BioProject accession number [PRJNA931302](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA931302).

ACKNOWLEDGMENTS

This work was conducted through the financial support of an NSF-HBCU-UP-RIA grant (number 1900164) and a subaward from Sandia National Labs Laboratory Directed Research and Development (LDRD) program. The work was also in part supported by the Norfolk State University Department of Biology. We thank David Gauthier (Old Dominion University) and the many members of the NSU undergraduate student body who participated in the genetics course-based undergraduate research experience and helped with the annotation of the Bassalto phage genome.

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC, under contract number DE-NA0003525 with the U.S. Department of Energy (DOE). This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the U.S. Government.

REFERENCES

- Doss J, Culbertson K, Hahn D, Camacho J, Barekzi N. 2017. A review of phage therapy against bacterial pathogens of aquatic and terrestrial organisms. *Viruses* 9:50. <https://doi.org/10.3390/v9030050>.
- 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>.
- Ecale Zhou CL, Malfatti S, Kimbrel J, Philipson C, McNair K, Hamilton T, Edwards R, Souza B. 2019. multiPhATE: bioinformatics pipeline for functional annotation of phage isolates. *Bioinformatics* 35:4402–4404. <https://doi.org/10.1093/bioinformatics/btz258>.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- McNair K, Zhou C, Dinsdale EA, Souza B, Edwards RA. 2019. PHANOTATE: a novel approach to gene identification in phage genomes. *Bioinformatics* 35:4537–4542. <https://doi.org/10.1093/bioinformatics/btz265>.
- 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).
- Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas 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>.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.955>.
- Pope WH, Carbonara ME, Cioffi HM, Cruz T, Dang BQ, Doyle AN, Fan OH, Gallagher M, Gentile GM, German BA, Farrell ME, Gerwig M, Hunter KL, Lefever VE, Marfisi NA, McDonnell JE, Monga JK, Quiroz KG, Pong AC, Rimple PA, Situ M, Sohnen PC, Stockinger AN, Thompson PK, Torchio NM, Toner CL, Ulbrich MC, Vohra NI, Zakir A, Adkins NL, Brown BR, Churilla BM, Kramer ZJ, Lapin JS, Montgomery MT, Prout AK, Grubb SR, Warner MH, Bowman CA, Russell DA, Hatfull GF. 2015. Genome sequences of mycobacteriophages AlanGrant, Baeae, Corofin, OrangeOswald, and Vincenzo, new members of cluster B. *Genome Announc* 3:e00586-15. <https://doi.org/10.1128/genomeA.00586-15>.
- Hatfull GF. 2012. The secret lives of mycobacteriophages. *Adv Virus Res* 82:179–288. <https://doi.org/10.1016/B978-0-12-394621-8.00015-7>.