



Complete Genome Sequence of *Gordonia rubripertincta* Bacteriophage Hexbug Suggests Potential for a New CT Subcluster

Daniel J. Barszczak,^a Maxwell J. Allison,^a Yaqi Cai,^a Lily Forman,^a Joshua M. Goldstein,^a Sarrah M. Hakimjee,^a Matthew T. Scapicchio,^a Ruben Torres,^a  Hannah E. Gavin^{b,c}

^aTufts University, Medford, Massachusetts, USA

^bExperimental College, Tufts University, Medford, Massachusetts, USA

^cDepartment of Biology, Tufts University, Medford, Massachusetts, USA

ABSTRACT Through the SEA-PHAGES program at Tufts University, a bacteriophage infecting *Gordonia rubripertincta* NRRL B-16540 was isolated and characterized. Hexbug is a lytic phage and is currently one of 44 phages belonging to cluster CT. The Hexbug genome shares >96% nucleotide identity with cluster CT phage Orla.

Bacteriophages (phages), which number approximately 10^{31} particles on Earth, are viruses that infect bacterial hosts (1, 2). Since the discovery of these entities in 1915, phages have shown potential to be used to treat problematic bacterial infections (3, 4). In collaboration with the SEA-PHAGES program, here we report the comprehensive genome sequence and annotation of a novel *Gordonia rubripertincta* phage, Hexbug (5, 6).

Hexbug was isolated from a moist garden soil sample collected in early September 2021 from Arlington, Massachusetts (42.412895 N, 71.13305 W). Briefly, the soil sample was washed with peptone-yeast extract-calcium (PYCa) medium, the wash was collected by centrifugation and filtration (0.02- μ m pore size), the filtrate was plated in top agar with *Gordonia rubripertincta* NRRL B-16540, and the plates were incubated at 30°C. Hexbug was purified by three successive rounds of single-plaque isolation. Negative-stain transmission electron microscopy revealed that Hexbug has siphovirus morphology (Fig. 1).

A high-titer lysate was generated from confluent plates. DNA was then extracted from the lysate with the Promega Wizard DNA clean-up system and prepared for sequencing with the NEBNext Ultra II fragmentation system (FS) prep kit. The library was sequenced on an Illumina MiSeq (v3 reagents) instrument at the University of Pittsburgh, yielding 48,574 150-bp reads with approximately $1,038\times$ coverage (7). The genome sequence was assembled *de novo* from raw reads with Newbler v2.9 and results checked with Consed v29 (8, 9). The Hexbug genome is 47,190 bp long, with a GC content of 63.3% and 10-bp 3' single-stranded overhangs (5'-CGGTAGGCAT-3').

Open reading frames (ORFs) in the Hexbug genome were initially autoannotated using GeneMark v4.9, Glimmer v3.02b, and ARAGORN v1.1 within DNA Master v5.0.2 (10–13). Gene starts were then refined using BLASTn searches against the PhagesDB Actinobacteriophage database and NCBI BLASTn standard database nucleotide collection. Information on coding potential, start site prevalence, and Shine Dalgarno sequence potential were gathered from GeneMark, Starterator, and DNAMaster, respectively (10, 13–16). Gene functions were predicted using Phamerator, NCBI BLASTp against standard databases, and HHPred against the following databases: PDB_mmCIF70_12_Oct, Pfam-A_v35, NCBI_Conserved_Domains(CD)_v3.18, and PRD_v6.9 (14, 17, 18). The genome was screened for antibiotic resistance genes on the CARD v3.2.0 RGI v5.2.1 Web portal (19). ORFs that could not be assigned a function were evaluated for transmembrane potential with TMHMM v2.0 and SOSUI (20–22).

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2022 Barszczak 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 Hannah E. Gavin, hannah.gavin@tufts.edu.

The authors declare no conflict of interest.

Received 13 September 2022

Accepted 8 October 2022

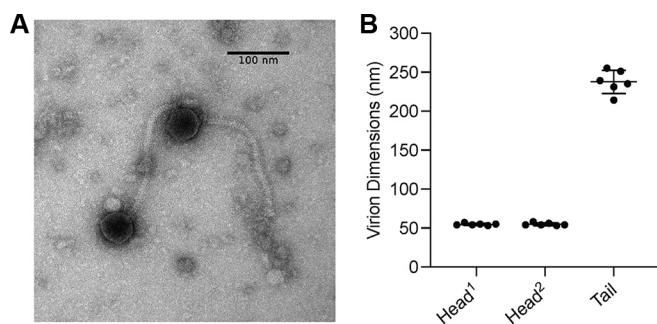


FIG 1 (A) Hexbug virions negatively stained with 1% uranyl acetate and visualized with transmission electron microscopy (TEM) at the Harvard Medical School Electron Microscopy Facility. Hexbug is a siphovirus with an isometric icosahedral head and long noncontractile tail (26). (B) Dimensions of 6 virions measured using ImageJ (27). Graph displays individual measurements as dots, with bars for mean and standard deviation. Scale bar, 100 nm. Hexbug tails measure $238 \text{ nm} \pm 15 \text{ nm}$ in length. Capsids are $55 \text{ nm} \pm 1 \text{ nm}$ along the axis perpendicular to the tail (Head¹) and $55 \text{ nm} \pm 2 \text{ nm}$ along the same axis as the tail (Head²).

Annotation identified 73 open reading frames, of which 36 were assigned functions. Genes connected to virion structure and assembly ($n = 16$) and genes facilitating cell lysis ($n = 3$) are situated in the left half of the genome. ORFs connected to DNA metabolism, including helicase and primase genes, are distributed across the right side of the genome. Five additional ORFs possess predicted transmembrane domains.

Actinobacteriophages that share at least 35% gene content similarity (GCS) are grouped into clusters; members often infect related hosts and show similar viral replication modes (4, 7, 23, 24). Hexbug is assigned to cluster CT according to the GCS analysis reflected in the Actinobacteriophage database as of 31 May 2022. Like other CT phages, Hexbug produces clear plaques and lacks lysogeny-associated genes, suggesting an obligately lytic replication cycle (7). The isometric capsid shape of Hexbug (Fig. 1) is also common in cluster CT.

Within cluster CT, Hexbug is most similar to phage Orla (GenBank accession [MN89453](#)), sharing 97% nucleotide identity across 99% of the genome. At >93%, the GCS between Hexbug and Orla is greater than their relationship to any other cluster CT phages and is comparable to GCS within phage subclusters (23). The relationship between these two phages foreshadows the emergence of the first phage subcluster within cluster CT. In application, lytic phages like Hexbug expand the arsenal of putative therapeutics for rare *Gordonia rubripertincta* infections in humans (25).

Data availability. The complete Hexbug genome is available at GenBank accession number [ON970609](#). The raw data can be accessed using Sequence Read Archive (SRA) number [SRX14443511](#).

ACKNOWLEDGMENTS

This work was made possible by the Howard Hughes Medical Institute SEA-PHAGES program and the joint efforts of the Tufts Experimental College and Department of Biology. Genome sequencing was performed at the University of Pittsburgh. Electron microscopy imaging, consultation, and services were performed in the Electron Microscopy Facility at Harvard Medical School.

We extend special thanks to Becky Garlena, Dan Russell, Graham Hatfull, and Deborah Jacobs-Sera for technical support. We also gratefully acknowledge Tufts Dean of Academic affair, Sam Thomas, for financial support; Howard Woolf and Amy Goldstein for infrastructural and financial support; Michael Grossi and Brendan Carson for direct support in the classroom; and Anja Nordstrom and Maria Ericsson for microscopy sample preparation and imaging.

REFERENCES

1. Comeau AM, Hatfull GF, Krisch HM, Lindell D, Mann NH, Prangishvili D. 2008. Exploring the prokaryotic virosphere. *Res Microbiol* 159:306–313. <https://doi.org/10.1016/j.resmic.2008.05.001>.
2. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* 96:2192–2197. <https://doi.org/10.1073/pnas.96.5.2192>.
3. Twort FW. 1915. An investigation on the nature of ultra-microscopic viruses. *The Lancet* 186:1241–1243. [https://doi.org/10.1016/S0140-6736\(01\)20383-3](https://doi.org/10.1016/S0140-6736(01)20383-3).
4. Hatfull GF. 2020. Actinobacteriophages: genomics, dynamics, and applications. *Annu Rev Virol* 7:37–61. <https://doi.org/10.1146/annurev-virology-122019-070009>.
5. 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. <https://doi.org/10.1128/mBio.01051-13>.
6. Hatfull GF. 2015. Innovations in undergraduate science education: going viral. *J Virol* 89:8111–8113. <https://doi.org/10.1128/JVI.03003-14>.
 7. Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. *Bioinformatics* 33:784–786. <https://doi.org/10.1093/bioinformatics/btw711>.
 8. 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>.
 9. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. *Bioinformatics* 29:2936–2937. <https://doi.org/10.1093/bioinformatics/btt515>.
 10. 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>.
 11. 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>.
 12. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
 13. Pope WH, Jacobs-Sera D. 2018. Annotation of bacteriophage genome sequences using DNA Master: an overview, p 217–229. *In* Clokie MRJ, Kropinski AM, Lavigne R (ed), *Bacteriophages: methods and protocols*, vol 3. Springer, New York, NY.
 14. 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).
 15. WUSTL. 2022. Starterator.
 16. Welkin H, Pope DJ-S, Russell DA, Cresawn SG, Hatfull GF. 2017. SEA-PHAGES Bioinformatics Guide, on University of Pittsburgh and Howard Hughes Medical Institute. <https://seaphagesbioinformatics.helpdocsonline.com/home>.
 17. Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248. <https://doi.org/10.1093/nar/gki408>.
 18. 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>.
 19. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>.
 20. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
 21. Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14:378–379. <https://doi.org/10.1093/bioinformatics/14.4.378>.
 22. Mitaku S, Hirokawa T, Tsuji T. 2002. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics* 18:608–616. <https://doi.org/10.1093/bioinformatics/18.4.608>.
 23. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, Russell DA, Warner MH, Hatfull GF, Science Education Alliance-Phage Hunters Advancing G, Evolutionary S. 2017. Bacteriophages of *Gordonia* spp. display a spectrum of diversity and genetic relationships. *mBio* 8:e01069-17. <https://doi.org/10.1128/mBio.01069-17>.
 24. Mavrich TN, Hatfull GF. 2017. Bacteriophage evolution differs by host, lifestyle and genome. *Nat Microbiol* 2:17112. <https://doi.org/10.1038/nmicrobiol.2017.112>.
 25. Ramanan P, Deziel PJ, Wengenack NL. 2013. *Gordonia* bacteremia. *J Clin Microbiol* 51:3443–3447. <https://doi.org/10.1128/JCM.01449-13>.
 26. Ackermann HW. 1998. Tailed bacteriophages: the order Caudovirales. *Adv Virus Res* 51:135–201. [https://doi.org/10.1016/S0065-3527\(08\)60785-x](https://doi.org/10.1016/S0065-3527(08)60785-x).
 27. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671–675. <https://doi.org/10.1038/nmeth.2089>.