




Genome Sequence of CaiB, a DR Cluster Actinobacteriophage That Infects *Gordonia rubripertincta*

Bienna Welsh,^a Nader M. Abdalla,^a Esteban Aldana,^a Veronica M. Alvarado Fernandez,^a Bruna Arenales Salgado de Oliveira,^a Diane Fakhre,^a Amelia J. Haymond,^a Katelyn M. Helton,^a Aditi Kanchibhatta,^a Jahwanza Knight,^a Sydney Marshall,^a Maomi Laine N. Martinez,^a Arielle Merkher,^a Savannah E. Morrow,^a Katie P. Nguyen,^a Jahanvi J. Patel,^a Somesh R. Patel,^a Pravalika Rayala,^a Kira M. Ruiz-Houston,^a Aarya P. Satardekar,^a Shifa M. Shaikh,^a Adrian E. Terron Osorio,^a Rachel C. Weitz,^a Louis Otero,^a  Richard S. Pollenz^a

^aDepartment of Cell Biology, Microbiology, and Molecular Biology, University of South Florida, Tampa, Florida, USA

ABSTRACT CaiB is a DR cluster actinobacteriophage that was isolated from soil in Florida using *Gordonia rubripertincta* NRRL B-16540 as the host. The genome is 61,620 bp, has a GC content of 68.6%, and contains 85 predicted protein coding genes. CaiB has several putative operons and has repeated intergenic regions that may be involved in gene regulation.

There are an estimated 10^{31} phage particles in the world, and 4,000 actinobacteriophage genomes have been annotated (1). Since approximately 70% of annotated protein coding genes do not have known function (2), the isolation of evolutionarily diverse actinobacteriophages helps advance the understanding of phage genomics and evolutionary science.

CaiB was isolated from a moist soil sample from Tampa, Florida (28.086388N, 82.384166W), using *Gordonia rubripertincta* NRRL B-16540 as the host. Bacterial infections were performed at 30°C utilizing peptone-yeast calcium agar (PYCa). Genomic DNA was isolated after three rounds of plaque purification using the Wizard DNA cleanup kit (A7280; Promega). Genomic DNA was used to create sequencing libraries with NEBNext Ultra II library preparation kit v3 reagents. Sequencing was performed by the Pittsburgh Bacteriophage Institute, and the libraries were run on an Illumina MiSeq instrument, yielding 295,653 paired-end 150-base reads with 625-fold average coverage. Raw reads were assembled with Newbler v2.9 (3), yielding a single phage contig. The results were checked for completeness, accuracy, and genome termini using Consed (4). Default parameters were used for all software unless otherwise specified. CaiB is circularly permuted based on a lack of defined genome ends (5) and was bioinformatically linearized such that base 1 is assigned in accord with other *Gordonia* phages (5). CaiB was autoannotated using DNA Master v5.23.6 (6), and all of the genes were then manually validated for correct starts and functional calls. GeneMark v2.5 (7) and Glimmer v3.02 (8) were utilized to assess start sites and coding potential, and Starterator v1.2 (2) was used to summarize the starts across each family of phage genes. To collect evidence for gene function and the validity of each gene product, HHpred v3.2 (9), NCBI BLAST (10), the Conserved Domain Database (CDD) (11), TMHMM v2.0 (12), and SOSUI (13) were utilized. tRNAscan-SE v2.0 (14) and ARAGORN v1.2.41 (15) were utilized to identify putative tRNAs and transfer-messenger RNAs.

Negative-staining transmission electron microscopy shows that CaiB has a 310-nm tail and icosahedral capsid of 62 nm (Fig. 1). CaiB has a 61,620-bp genome, has a GC content of 68.6%, and contains 85 predicted protein coding genes. CaiB is one of 11 phages in the DR cluster (Table 1). DR phages have similar genome organizations and morphologies. The CaiB genome shows between 65 and 97% nucleotide identity to the other DR members (Table 1) (16). Genes 1 to 5 precede the terminase, have 4-bp overlaps, and encode a series of nucleotide modification enzymes, including

Editor John J. Dennehy, Queens College CUNY

Copyright © 2022 Welsh 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 Richard S. Pollenz, pollenz@usf.edu.

The authors declare no conflict of interest.

Received 21 April 2022

Accepted 4 June 2022

Published 27 June 2022

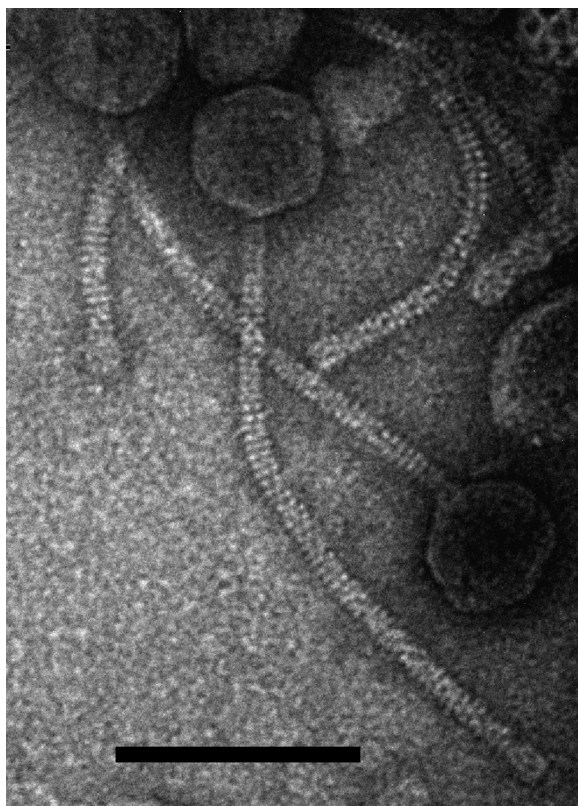


FIG 1 Transmission electron micrograph of *Gordonia* phage CaiB (<https://phagesdb.org/phages/CaiB>). Phage lysates were negatively stained with 1% uranyl acetate. Scale bar = 120 nm.

ParB-like nuclease, TET/JBP oxidoreductase, pyrophosphorylase, and adenylate kinase, which may be involved in evasion of host restriction systems (17). CaiB contains several repeated intergenic regions upstream of genes 42, 71, 77, 80, 81, and 85. These regions contain 100% identity in positions -4 to -11 (5'-GAGAGGAC-3'), -14 to -21 (5'-ACCCGCTC-3'), -25 to -30 (5'-GCGGGA-3'), and -35 to -40 (5'-TGTTGT-3') and may serve a function in regulation.

Data availability. This whole-genome shotgun project has been deposited in DDB/ENA/GenBank under the accession numbers [ON108644](https://www.ncbi.nlm.nih.gov/nuccore/ON108644) and [SRX14597705](https://www.ncbi.nlm.nih.gov/nuccore/SRX14597705). The version described in this paper is the first version. Data for CaiB are archived in Phamerator (18) and the Actinobacteriophage Database at PhagesDB.org (2) (<https://phagesdb.org/phages/CaiB>).

TABLE 1 DR cluster phage characteristics and nucleotide identity values

Phage	GenBank accession no.	Genome size (bp)	GC content (%)	Nucleotide identity (%) with:											
				NHagos	CaiB	AnarQue	CloverMinnie	Mariokart	Axumite	Ligma	Sour	BiggityBass	AnClar	Yago84	
NHagos	MN369758	59,580	68.2	100											
CaiB	ON108644	61,620	68.6	75.88	100										
AnarQue	OK216879	61,822	68.8	76.19	96.96	100									
CloverMinnie	MN234196	61,098	68.7	76.23	96.94	97.98	100								
Mariokart	MT657335	60,762	70.5	67.29	66.44	66.72	66.81	100							
Axumite	ON081333	61,714	70.2	67.44	65.99	66.31	66.43	81.79	100						
Ligma	OM105886	61,714	70.2	67.44	65.99	66.31	66.43	81.79	99.99	100					
Sour	MH153810	61,670	68.0	64.11	64.70	64.86	64.97	71.01	71.62	71.62	100				
BiggityBass	ON260813	63,202	69.4	66.16	65.23	65.30	65.43	73.47	74.76	74.76	72.74	100			
AnClar	MN908693	61,856	69.8	66.03	65.14	65.22	65.29	73.31	74.75	74.76	73.04	88.56	100		
Yago84	MK801725	61,890	70.0	66.00	65.18	65.30	65.37	73.06	74.55	74.55	73.97	89.30	98.49	100	

ACKNOWLEDGMENTS

This work was made possible by the Howard Hughes Medical Institute (HHMI) Science Education Alliance and the University of Pittsburgh.

R. A. Garlena, D. A. Russell, D. Jacobs-Sera, and G. H. Hatfull are gratefully acknowledged for their help with genome sequencing and quality control, and we also acknowledge support of the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program (19).

REFERENCES

- Rohwer F, Youle M, Maughan H, Hisakawa N. 2014. Life in our phage world. a centennial field guide to the Earth's most diverse inhabitants. Wholon, San Diego, CA.
- Russell DA, Hatfull GF. 2017. PhagesDB: the Actinobacteriophage Database. *Bioinformatics* 33:784–786. <https://doi.org/10.1093/bioinformatics/btw711>.
- Genivaldo GZ, Silva BE, Dutilh D, Matthews K, Elkins R, Schmieder EA, Dinsdale RAE. 2013. Combining de novo and reference-guided assembly with scaffold_builder. *Source Code Biol Med* 8:23. <https://doi.org/10.1186/1751-0473-8-23>.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical toll of sequence finishing. *Genome Res* 8:195–202. <https://doi.org/10.1101/gr.8.3.195>.
- Russel 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.
- Pope WH, Jacobs-Sera D. 2018. Annotation of bacteriophage genome sequences using DNA Master: an overview. *Methods Mol Biol* 1681:217–229. https://doi.org/10.1007/978-1-4939-7343-9_16.
- Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes, and viruses. *Nucleic Acids Res* 33:451–454. <https://doi.org/10.1093/nar/gki487>.
- 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>.
- Söding 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>.
- 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).
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's Conserved Domain Database. *Nucleic Acids Res* 43:D222–D226. <https://doi.org/10.1093/nar/gku1221>.
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
- 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>.
- Lowe T, Chan P. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57. <https://doi.org/10.1093/nar/gkw413>.
- Laslett D, Canback B. 2004. ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
- Lee I, Ouk Kim Y, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100–1103. <https://doi.org/10.1099/ijsem.0.000760>.
- Iyer LM, Zhang D, Burroughs AM, Aravind L. 2013. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res* 41:7635–7655. <https://doi.org/10.1093/nar/gkt573>.
- 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>.
- Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, 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 under-graduate students. *mBio* 5:e01051-13. <https://doi.org/10.1128/mBio.01051-13>.