




# Assembly and Annotation of the Complete Genome Sequence of T4-Like Bacteriophage 132

 C. Roush,<sup>a,b</sup>  B. K. Chan,<sup>a</sup>  P. E. Turner,<sup>a,b,c</sup>  A. R. Burmeister<sup>a,b</sup>

<sup>a</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA

<sup>b</sup>BEACON Center for the Study of Evolution in Action, East Lansing, Michigan, USA

<sup>c</sup>Program in Microbiology, Yale School of Medicine, New Haven, Connecticut, USA

**ABSTRACT** Here, we describe the complete sequence of bacteriophage 132, a T4-like *Escherichia coli* phage. Phage 132 has a genome of 166,922-bp length, with 286 predicted genes.

**B**acteriophage 132 was isolated from feces from a swine farm in Connecticut, USA. Like phage T4, it is reliant on host lipopolysaccharide (LPS) synthesis genes (1); however, its outer membrane protein receptor is likely OmpF (1), rather than OmpC (the T4 receptor) (2).

Swine feces were resuspended in phosphate-buffered saline with 10 mM MgSO<sub>4</sub>. This mixture was centrifuged at 3,220 × *g* for 15 min and filtered through a 0.22- $\mu$ m filter. The filtrate was enriched in tryptic soy broth and *Escherichia coli* strain BW25113 for 24 h. Plaques were triple isolated and amplified to high titer. DNA was extracted using the Norgen Biotek phage DNA isolation kit. The sequencing library was prepared using the Nextera DNA Flex library preparation kit and sequenced at Felix Biotechnology on an Illumina NextSeq system, yielding 150-bp paired-end reads. Only the R1 reads were used for annotation. Genome annotation and assembly were conducted in the Galaxy (3) and Web Apollo (4) phage annotation platforms. Default parameters were used for all software unless otherwise noted. Sequencing reads were rarified to 250× coverage to improve assembly using FastQ Subset v.1.1 (5, 6) and trimmed using the Trim Sequences tool v.1.0.2 (7). Reads were quality checked using FastQC v.0.72 (8), and a contig was assembled at 128× coverage using SPAdes v.3.12.0 (9). One other contig, a 56-bp poly (G) artifact from sequencing (10), was omitted from further analysis. One of the overlapping ends (55 bp), which indicated a circular genome, was trimmed using the Trim Sequences tool v.1.0.2 (7).

Using BLASTn (11) with the complete genome, phage 132 was found to be closely related to the well-characterized phage T4 (GenBank accession number [MT984581](https://www.ncbi.nlm.nih.gov/nuclseq/MT984581)), with 95.36% nucleotide identity over 88% of the phage 132 genome. The complete assembled contig was reopened in alignment with the T4 reference genome using Reopen Fasta Sequences v.2.0 (5, 6). This sequence was imported into Apollo using the Galaxy Structural Phage Annotation Workflow v.2020.07 (6). Gene locations were predicted by weighing evidence from the algorithms GLIMMER3 v.0.2 (12), MetaGeneAnnotator v.1.0.0 (13), and SixPack v.5.0.0 (14). Final protein-coding gene calls were made manually, based on the presence of a valid start codon, the presence of a valid Shine-Dalgarno sequence, and assessment of gaps between genes. Calls for tRNAs were made by examining evidence from tRNAScan-SE v.0.4 (15) and ARAGORN v.0.6 (16). One predicted gene spanned the ends of the genome; therefore, we adjusted the genome start location using Reopen Fasta Sequences v.2.0 (5, 6) to keep the ends of that gene together. Functional predictions were made manually using the Galaxy Functional Phage Annotation Workflow v.2020.09 (6) and comparing BLASTp (17, 18) results from three phage genome databases (the canonical

**Citation** Roush C, Chan BK, Turner PE, Burmeister AR. 2021. Assembly and annotation of the complete genome sequence of T4-like bacteriophage 132. *Microbiol Resour Announc* 10:e00649-21. <https://doi.org/10.1128/MRA.00649-21>.

**Editor** John J. Dennehy, Queens College CUNY

**Copyright** © 2021 Roush 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 A. R. Burmeister, [alita.burmeister@yale.edu](mailto:alita.burmeister@yale.edu).

**Received** 14 July 2021

**Accepted** 1 September 2021

**Published** 30 September 2021

phages, nonredundant-all phages, and Swiss-Prot databases). The geecee tool v.5.0.0 (19) was used to determine that the GC content of this genome is 35%.

The phage 132 genome contains 286 predicted genes, 150 of which are annotated as hypothetical proteins. A total of 126 predicted genes are annotated as putative proteins with predicted functions, and 10 tRNA genes are identified.

**Data availability.** The annotated phage 132 genome is available in GenBank under accession number [MZ417521](https://doi.org/10.1093/nar/gkx1095). The rarified sequence reads are available in the Sequence Read Archive (SRA) with SRA accession number [SRP325389](https://doi.org/10.1093/nar/gkx1095). The BioProject accession number is [PRJNA740247](https://doi.org/10.1093/nar/gkx1095).

## ACKNOWLEDGMENTS

This work was supported by NIH grant R21AI144345 from the National Institute of Allergy and Infectious Diseases to A.R.B. and P.E.T., by the BEACON Center for the Study of Evolution in Action (NSF cooperative agreement DBI-0939454), and by a STARS II fellowship from Yale University to C.R. Sequencing was provided by Felix Biotechnology, a company that seeks to commercially develop phages for use as therapeutics.

## REFERENCES

- Burmeister AR, Bender RG, Fortier A, Lessing AJ, Chan BK, Turner PE. 2018. Two lytic bacteriophages that depend on the *Escherichia coli* multi-drug efflux gene *toIC* and differentially affect bacterial growth and selection. *bioRxiv* 397695. <https://doi.org/10.1101/397695>.
- Yu F, Mizushima S. 1982. Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. *J Bacteriol* 151:718–722. <https://doi.org/10.1128/jb.151.2.718-722.1982>.
- Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46:W537–W544. <https://doi.org/10.1093/nar/gky379>.
- Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elisk CG, Lewis SE. 2013. Web Apollo: a web-based genomic annotation editing platform. *Genome Biol* 14:R93. <https://doi.org/10.1186/gb-2013-14-8-r93>.
- Mijalis E, Rasche H. 2017. CPT Galaxy tools. <https://github.com/tamu-cpt/galaxy-tools>.
- Ramsey J, Rasche H, Maughmer C, Criscione A, Mijalis E, Liu M, Hu JC, Young R, Gill JJ. 2020. Galaxy and Apollo as a biologist-friendly interface for high-quality cooperative phage genome annotation. *PLoS Comput Biol* 16:e1008214. <https://doi.org/10.1371/journal.pcbi.1008214>.
- Gordon A. 2010. FASTQ/A short-reads pre-processing tools. [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit).
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Stoler N, Nekrutenko A. 2021. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom Bioinform* 3:lqab019. <https://doi.org/10.1093/nargab/lqab019>.
- NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 46:D8–D13. <https://doi.org/10.1093/nar/gkx1095>.
- 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>.
- Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* 15:387–396. <https://doi.org/10.1093/dnares/dsn027>.
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47:W636–W641. <https://doi.org/10.1093/nar/gkz268>.
- Lowe TM, Chan PP. 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 to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N. 2015. NCBI BLAST+ integrated into Galaxy. *Gigascience* 4:39. <https://doi.org/10.1186/s13742-015-0080-7>.
- Blankenberg D, Taylor J, Schenck I, He J, Zhang Y, Ghent M, Veeraraghavan N, Albert I, Miller W, Makova KD, Hardison RC, Nekrutenko A. 2007. A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. *Genome Res* 17:960–964. <https://doi.org/10.1101/gr.5578007>.