





# Complete Genome Sequence of a *Bacteroides fragilis* Bacteriophage, vB\_BfrS\_NCTC

 Mohammad A. Tariq,<sup>a</sup>  Simon R. Carding<sup>a,b</sup>

<sup>a</sup>Gut Microbes and Health Research Programme, Quadram Institute Bioscience, Norwich, United Kingdom

<sup>b</sup>Norwich Medical School, University of East Anglia, Norwich, United Kingdom

**ABSTRACT** *Bacteroides fragilis* is an obligate anaerobe and a common gut commensal bacterium that is also an important opportunistic pathogen. Here, we present the complete genome sequence of the circularly permuted *B. fragilis* bacteriophage vB\_BfrS\_NCTC. It comprises 47,160 bp, with 69 open reading frames.

**B** *Bacteroides fragilis* is a constituent of the human colonic microbiota that, as a result of disruptions due to, for example, infections or antibiotic exposure, can gain access to the systemic circulation, leading to severe and sometimes fatal disease (1). It is inherently resistant to many antibiotics, and the rise in antimicrobial resistance has led to a renewed interest in phage therapy (2). We have isolated a novel bacteriophage, vB\_BfrS\_NCTC, and provide its complete genome sequence here.

The bacteriophage was isolated from samples obtained from a sewage wastewater treatment facility in the United Kingdom; sewage filtrate was screened against *B. fragilis* NCTC 9343. A 100-ml aliquot of wastewater was filtered through a 0.2- $\mu$ m polyethersulfone (PES) syringe filter (Minisart; Sartorius), and the filtrate was concentrated using a 100-kDa cellulose centrifugal filter (Amicon Ultra-15; Millipore). The concentrated sample was resuspended in phosphate-buffered saline (PBS) to 1,500  $\mu$ l. This sewage concentrate was added to 5 ml semisoft agar (0.35% [wt/vol]) containing 200  $\mu$ l *B. fragilis* NCTC 9343 in mid-exponential growth phase. The plate was incubated for 16 h in an anaerobic cabinet, and resulting single plaques were purified an additional three times before propagation using the *Bacteroides* phage recovery medium (BPRM) top agar method (3). In brief, purified phage lysate was mixed with 200  $\mu$ l mid-log-growth-phase bacterial culture in semisoft BPRM agar (0.35% [wt/vol]) and incubated for 16 h in an anaerobic cabinet (5% CO<sub>2</sub>, 5% H<sub>2</sub>, and 90% N<sub>2</sub> at 37°C and  $\sim$ 25 lb/in<sup>2</sup>). SM buffer was used to harvest the lawn until the phage titer reached  $>10^9$  PFU/ml. The phage DNA was extracted using a phage kit (catalog number 46800; Norgen Biotek Corp.) according to the manufacturer's instructions. The DNA was sequenced by the Quadram Institute Bioscience sequencing facility (Norwich, UK) on an Illumina NextSeq 500 system using the Nextera XT library preparation kit (2  $\times$  150 cycles, v2 chemistry; Illumina, Saffron Walden, UK). For all further steps, standard settings were used unless stated otherwise. Paired-end sequencing reads were provided as FASTQ files. Adapters were removed using Trimmomatic v0.39 (4), and quality trimming was carried out using Sickle v1.33 at  $-q$  30 and  $-l$  15 (5). The read quality was assessed using FastQC v0.11.9 (6). These reads were *de novo* assembled using SPAdes v3.14.1 (7). The assembled contigs were determined to be circularly permuted and were confirmed with Circlator v1.5.5 using the Minimus2 circularization pipeline as an option (8). The reads were mapped back using BWA-MEM v0.7.17 (9) and SAMtools v1.12 (10), and coverage statistics were generated using Qualimap v2.2.1 (11). The genome open reading frames (ORFs) were predicted and initially annotated using the RAST server (12, 13). To comprehensively annotate the genome, ORFs were searched with BLASTp (14) against the NCBI nonredundant database (access date: 15 November 2020) at an E value of  $<1 \text{ e}^{-5}$ . ARAGORN was used to identify any tRNA genes (15).

**Citation** Tariq MA, Carding SR. 2021. Complete genome sequence of a *Bacteroides fragilis* bacteriophage, vB\_BfrS\_NCTC. Microbiol Resour Announc 10:e00548-21. <https://doi.org/10.1128/MRA.00548-21>.

**Editor** David Rasko, University of Maryland School of Medicine

**Copyright** © 2021 Tariq and Carding. 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 Mohammad A. Tariq, [mohammad.tariq@quadram.ac.uk](mailto:mohammad.tariq@quadram.ac.uk).

**Received** 27 May 2021

**Accepted** 21 June 2021

**Published** 22 July 2021

The sequencing of vB\_BfrS\_NCTC yielded 1,012,042 clean reads, with a mean length of 129 bp. The reads had an average Phred quality score of 35, and the GC content was 39.56%. A total of 98.54% of the reads mapped back to the genome, giving a mean coverage depth of 2,793 $\times$ . The assembled genome was shown to be circularly permuted at 47,160 bp, with a GC content of 38.83%. A total of 69 ORFs, with no tRNA genes, were identified. The genome comparison was performed using the default settings of the online tool BLASTN with the nonredundant/nucleotide database (16) and showed the greatest similarity at the nucleotide level to the *B. fragilis* phage Barc2635 (GenBank accession number MN078104), with 92% query coverage and 97% identity.

**Data availability.** The genome sequence of vB\_BfrS\_NCTC is available in GenBank under the accession number MW314138. The SRA data can be found under the accession number SRR13174062.

## ACKNOWLEDGMENTS

This work was supported in part by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and BBSRC Institute Strategic Program grant BB/R012490/1 to the Gut Microbes and Health Research Programme and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356 410 (S.R.C.).

We thank Dave Baker and Steven Rudder for sequencing this phage at Quadram Institute Bioscience (Norwich, UK) during the coronavirus disease 2019 (COVID-19) pandemic.

## REFERENCES

- Niestępski S, Harnisz M, Korzeniewska E, Aguilera-Arreola MG, Contreras-Rodríguez A, Filipkowska Z, Osińska A. 2019. The emergence of antimicrobial resistance in environmental strains of the *Bacteroides fragilis* group. *Environ Int* 124:408–419. <https://doi.org/10.1016/j.envint.2018.12.056>.
- Bakuradze N, Makalatia K, Merabishvili M, Togoshvili L, Chanishvili N. 2018. Selection of the active phages against *B. fragilis* for further study of therapeutic perspectives. *Georgian Med News* 111–116.
- Tariq MA, Newberry F, Haagmans R, Booth C, Wileman T, Hoyles L, Clokie MRJ, Ebdon J, Carding SR. 2020. Genome characterization of a novel wastewater *Bacteroides fragilis* bacteriophage (vB\_BfrS\_23) and its host GB124. *Front Microbiol* 11:583378. <https://doi.org/10.3389/fmicb.2020.583378>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
- Andrews S, Krueger F, Segonds-Pchon A, Biggins L, Krueger C, Wingett S, Montgomery J. 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>.
- Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997v2. <https://arxiv.org/abs/1303.3997>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Okonechnikov K, Conesa A, Garcia-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294. <https://doi.org/10.1093/bioinformatics/btv566>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia FF. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
- 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).
- 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>.
- NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 44:D7–D19. <https://doi.org/10.1093/nar/gkv1290>.