



Complete Genome Sequence of *Paraclostridium bifermentans* DSM 14991

 G. T. Little,^a L. Sellés Vidal,^{b,c} M. Steadman,^{b,c} R. Leyden-Preece,^{a,b,c} G. M. Taylor,^{b,c}  J. T. Heap^{a,b,c}

^aSchool of Life Sciences, The University of Nottingham, University Park, Nottingham, United Kingdom

^bDepartment of Life Sciences, Imperial College London, London, United Kingdom

^cImperial College Centre for Synthetic Biology, Imperial College London, London, United Kingdom

ABSTRACT The complete genome sequence of *Paraclostridium bifermentans* was obtained by assembly of Illumina and Oxford Nanopore (ONT) reads. The sequence will enable study into the organism's ability to biohydrogenate unsaturated acyl chains in the transformation of C20 polyunsaturated fatty acids (PUFAs) into the corresponding bioactive non-methylene-interrupted fatty acids (NMIFAs).

Paraclostridium bifermentans is a rod shaped, endospore forming, Gram-positive, obligately anaerobic, non-pathogenic bacterium (1, 2). Isolates of *Paraclostridium bifermentans* from areas where anophelines are endemic encode PMP1, a clostridial-like neurotoxin that selectively targets anopheline mosquitoes (3). The organism is capable of the unusual biohydrogenation of unsaturated acyl chains, transforming C20 polyunsaturated fatty acids (PUFAs) into the corresponding non-methylene-interrupted fatty acids (NMIFAs) (4), which may have novel bioactive properties (5).

Paraclostridium bifermentans type strain DSM 14991 (ATCC 638; JCM 1386; <https://www.dsmz.de/collection/catalogue/details/culture/DSM-14991>) was purchased from DSMZ. Spores were plated anaerobically at 37°C onto a Gifu anaerobic medium (GAM) agar plate. A single colony was grown anaerobically overnight in 10 mL GAM media, and genomic DNA was obtained by phenol-chloroform-isoamyl alcohol extraction (6). All library preparation and sequencing steps were performed by MicrobesNG (Birmingham, UK) from the same genomic DNA sample. An Illumina library was prepared using the Nextera XT library prep kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: input DNA increased 2-fold and PCR elongation time increased to 45 s. Pooled libraries were quantified using the Kapa Biosystems library quantification kit for Illumina and sequenced using a HiSeq 2500 instrument with a 250-bp paired-end protocol. Reads were adapter trimmed using Trimmomatic 0.30 (7) with a sliding window quality cutoff of Q15, to give 847,664 paired Illumina reads. Quality control (QC) reports for the Illumina data were generated using FastQC 0.11.9 (8) and aggregated using MultiQC 1.10.1 (9). Default parameters were used for all software unless otherwise specified.

A long read genomic DNA library was prepared with the Oxford Nanopore SQK-RBK004 kit using 400 to 500 ng of high-molecular-weight DNA. Barcoded samples were pooled into a single sequencing library and loaded in a FLO-MIN106 (R.9.4.1) flow cell in a GridION system (Oxford Nanopore [ONT], United Kingdom). Guppy 3.0.6 (10) was used for base calling. The ONT data were subject to QC using NanoPlot 1.38.0 (11), with mean read length, 3,770 bp; median read length, 2,384 bp; number of reads, 22,812; and read length N_{50} , 6,390 bp.

The genome sequence was assembled using Unicycler 0.4.8 (12) with parameters `min_fasta_length 2000` and bold mode, with the chromosome split across 3 incomplete contigs due to incomplete long read coverage. To achieve a single circular contig, two 40 kbp sequences from *Clostridium bifermentans* strain cbm (accession number [NZ_CP032452.1](https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP032452.1); 2067654..2107654, 2123574..2163574) were added as long read scaffolds. All scaffold sequences were updated

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2022 Little 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 J. T. Heap, john.heap@nottingham.ac.uk.

The authors declare no conflict of interest.

Received 4 February 2022

Accepted 14 June 2022

Published 27 June 2022

TABLE 1 Length and NCBI accession numbers of the sequenced chromosome and plasmids

| Name | Length (bp) | NCBI accession |
|------------------|------------------------|----------------|
| Chromosome | 3,304,779 | CP079737.1 |
| Plasmid unnamed1 | 142,041 | CP079738.1 |
| Plasmid unnamed2 | 52,569 | CP079739.1 |
| Plasmid unnamed3 | 47,028 | CP079740.1 |
| Plasmid unnamed4 | 19,800 | CP079741.1 |
| Total | 3,566,217 ^a | |

^a GC content of 28.6%.

during the assembly by short-read Illumina data from DSM 14991. The assembly produced 5 complete circular contigs (Table 1). The assembly was subject to QC by visualization of the assembly graphs with Bandage 0.8.1 (13). Annotation was done using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) 5.2 (14).

Data availability. The nucleotide sequences of the *Paraclostridium bifermentans* DSM 14991 chromosome and plasmids have been deposited in NCBI GenBank under accession numbers CP079737 to CP079741. Sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRX11627627 and SRX11627626.

ACKNOWLEDGMENTS

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) Network in Industrial Biotechnology and Bioenergy (NIBB) Plants to Products (P2P) grant HD-RD0300L, Croda Europe Limited, and BBSRC NIBB High Value Biorenewables Network grant POC-HVB-2020/02.

Genome sequencing was provided by MicrobesNG which is supported by the BBSRC (grant number BB/L024209/1).

REFERENCES

- Tissier H, Martelly A. 1902. Recherches sur la putréfaction de la viande de boucherie. *Ann Inst Pasteur* 16:865–903.
- Brooks ME, Epps HB. 1959. Taxonomic studies of the genus *Clostridium*: *Clostridium bifermentans* and *C. sordellii*. *J Gen Microbiol* 21:144–155. <https://doi.org/10.1099/00221287-21-1-144>.
- Contreras E, Masuyer G, Qureshi N, Chawla S, Dhillon HS, Lee HL, Chen J, Stenmark P, Gill SS. 2019. A neurotoxin that specifically targets *Anopheles* mosquitoes. *Nat Commun* 10:2869. <https://doi.org/10.1038/s41467-019-10732-w>.
- Sakurama H, Kishino S, Mihara K, Ando A, Kita K, Takahashi S, Shimizu S, Ogawa J. 2014. Biohydrogenation of C20 polyunsaturated fatty acids by anaerobic bacteria. *J Lipid Res* 55:1855–1863. <https://doi.org/10.1194/jlr.M045450>.
- Lehnen TE, da Silva MR, Camacho A, Marcadenti A, Lehnen AM. 2015. A review on effects of conjugated linoleic fatty acid (CLA) upon body composition and energetic metabolism. *J Int Soc Sports Nutr* 12:36. <https://doi.org/10.1186/s12970-015-0097-4>.
- Cartman ST, Minton NP. 2010. A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol* 76:1103–1109. <https://doi.org/10.1128/AEM.02525-09>.
- 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>.
- Babraham bioinformatics. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
- Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20:129. <https://doi.org/10.1186/s13059-019-1727-y>.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.