



HAL
open science

Complete Genome Sequence of *Bradyrhizobium* sp. Strain BDV5040, Representative of Widespread Genospecies B in Australia

Christine Oger-Desfeux, Jérôme Briolay, Phil M. Oger, Bénédicte Lafay

► To cite this version:

Christine Oger-Desfeux, Jérôme Briolay, Phil M. Oger, Bénédicte Lafay. Complete Genome Sequence of *Bradyrhizobium* sp. Strain BDV5040, Representative of Widespread Genospecies B in Australia. *Microbiology Resource Announcements*, 2021, 10 (3), pp.e01326-20. 10.1128/MRA.01326-20. hal-03118246

HAL Id: hal-03118246

<https://hal.science/hal-03118246v1>


Submitted on 22 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Complete Genome Sequence of *Bradyrhizobium* sp. Strain BDV5040, Representative of Widespread Genospecies B in Australia

Christine Oger-Desfeux,^a Jérôme Briolay,^b Philippe M. Oger,^c  Bénédicte Lafay^{d,e}

^aUniversité de Lyon, Université Claude Bernard Lyon 1, PRABI-ASMB, FR3728 Bio-Environnement et Santé, Villeurbanne, France

^bUniversité de Lyon, Université Claude Bernard Lyon 1, CNRS, DTAMB, FR3728 Bio-Environnement et Santé, Villeurbanne, France

^cUniversité de Lyon, INSA de Lyon, CNRS, UMR5240, Villeurbanne, France

^dUniversité de Lyon, École Centrale de Lyon, CNRS, UMR5005, Laboratoire Ampère, Écully, France

^eUniversité de Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France

ABSTRACT We report the complete genome sequence of *Bradyrhizobium* sp. strain BDV5040, representative of *Bradyrhizobium* genospecies B, which symbiotically associates with legume hosts belonging to all three Fabaceae subfamilies across the Australian continent. The complete genome sequence provides a genetic reference for this Australian genospecies.

B*radyrhizobium* sp. strain BDV5040 was isolated in 1995 from a root nodule of *Bossiaea ensata* (Fabaceae, Faboideae, Bossiaeeae) collected in Ben Boyd National Park, New South Wales, Australia (37°12'S, 149°57'E; altitude, 140 m), in the course of a survey of rhizobia associated with native shrubby legumes in southeastern Australia (1). It is a representative of *Bradyrhizobium* genospecies B, which occurs under different climatic and edaphic conditions across the whole Australian continent and exhibits a broad host range encompassing all three Fabaceae subfamilies (1–4).

Strain BDV5040 was grown from a lyophilized stock in 30 ml of yeast extract mannitol broth (5) at 25°C and 200 rpm for 5 days. Genomic DNA was prepared by successive phenol-chloroform extractions as described (6). DNA quantification and quality control were performed using a NanoDrop spectrophotometer, a Qubit 4 fluorometer, and agarose gel electrophoresis. The same DNA was used for Nanopore and Illumina sequencing. Illumina libraries were obtained using the Nextera XT kit following the manufacturer's instructions, starting with 1 ng of genomic DNA, and were analyzed by paired-end 2 × 300-bp sequencing on a MiSeq instrument. Poor-quality regions (Q < 30) of raw reads were removed using Sickle v1.33 (7). Long reads were obtained with the Oxford Nanopore Technologies MinION FLO-MIN106 flow cell from a library prepared with the SQK-RAD004 kit using 1 μg of DNA and 20-second tagmentation. Base calling was performed using Guppy v3.1.5. Sequence quality was assessed using FastQC v0.11.9 (8) for Illumina reads and MinIONQC.R v1.4.2 (9) for Nanopore reads. The preprocessing of Illumina sequencing data using Trimmomatic v0.39 (10) (parameters: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50) resulted in 3.87 million paired-end reads (mean lengths of 222 bp and 160 bp for forward and reverse reads, respectively, with ~150× coverage). Nanopore long reads (>1,500 bp) and quality (score of >8) using Nanofilt v2.5.0 (11), and adapters were removed using Porechop v0.2.4 (12). Long reads were further reduced to 800 Mbp as a target quantity using Filtlong v0.2.0 (13) (parameters: --min_length 2000 --keep_percent 90 --target_bases 800000000). Illumina and Nanopore reads were co-assembled using Unicycler v0.4.8 (14) with default parameters, resulting in a single

Citation Oger-Desfeux C, Briolay J, Oger PM, Lafay B. 2021. Complete genome sequence of *Bradyrhizobium* sp. strain BDV5040, representative of widespread genospecies B in Australia. *Microbiol Resour Announc* 10:e01326-20. <https://doi.org/10.1128/MRA.01326-20>.

Editor David A. Baltus, University of Arizona

Copyright © 2021 Oger-Desfeux 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 Bénédicte Lafay, benedicte.lafay@univ-lyon1.fr.

Received 21 November 2020

Accepted 18 December 2020

Published 21 January 2021

component with eight segments and incomplete status (length, 7,622,333 bp; N_{50} , 7,339,313 bp). Completion was obtained by exporting the sequence path from Bandage v0.8.1 (15) and filling a last gap using Pilon v1.23 (16) and by manually comparing the sequence with Unicycler 003_long_read_assembly.fasta. The assembly and complete chromosome sequence were carefully inspected by visualizing the alignment of long and short reads using minimap2 v2.17 (17) and IGV v2.7.2 (18). Finally, the chromosome was rotated to start at *dnaA*.

The circular chromosome is 7,622,528 bp long, with an average G+C content of 63.92%. The sequence was automatically annotated by the NCBI Prokaryote Genome Annotation Pipeline (PGAP) v4.13 (19). The genome consists of 7,092 protein-coding genes, 48 tRNAs, 1 copy each of the 5S, 16S, and 23S rRNA genes, and 88 pseudogenes.

Data availability. The genome sequence of *Bradyrhizobium* genospecies B strain BDV5040 is available in NCBI GenBank under accession number [CP061379](https://doi.org/10.1093/bioinformatics/bty149). The raw sequence reads are available under SRA accession numbers [SRX9514896](https://doi.org/10.1093/bioinformatics/bty149) and [SRX9514898](https://doi.org/10.1093/bioinformatics/bty149) under BioProject number [PRJNA662585](https://doi.org/10.1093/bioinformatics/bty149) and BioSample number [SAMN16089659](https://doi.org/10.1093/bioinformatics/bty149).

ACKNOWLEDGMENT

This project was supported by the École Centrale de Lyon BQR program.

REFERENCES

- Lafay B, Burdon JJ. 1998. Molecular diversity of rhizobia occurring on native shrubby legumes in southeastern Australia. *Appl Environ Microbiol* 64:3989–3997. <https://doi.org/10.1128/AEM.64.10.3989-3997.1998>.
- Lafay B, Burdon JJ. 2001. Small-subunit rRNA genotyping of rhizobia nodulating Australian *Acacia* spp. *Appl Environ Microbiol* 67:396–402. <https://doi.org/10.1128/AEM.67.1.396-402.2001>.
- Lafay B, Burdon JJ. 2006. Molecular diversity of rhizobia nodulating the invasive legume *Cytisus scoparius* in Australia. *J Appl Microbiol* 100:1228–1238. <https://doi.org/10.1111/j.1365-2672.2006.02902.x>.
- Lafay B, Burdon JJ. 2007. Molecular diversity of legume root-nodule bacteria in Kakadu National Park, Northern Territory, Australia. *PLoS One* 2: e277. <https://doi.org/10.1371/journal.pone.0000277>.
- Vincent JM. 1970. A manual for the practical study of root-nodule bacteria. International Biological Programme handbook no. 15. Blackwell Science Publications, Oxford, England.
- Dhaese P, De Greve H, Decraemer H, Schell J, Van Montagu M. 1979. Rapid mapping of transposon insertion and deletion mutations in the large Ti-plasmids of *Agrobacterium tumefaciens*. *Nucleic Acids Res* 7:1837–1849. <https://doi.org/10.1093/nar/7.7.1837>.
- Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* 35:523–525. <https://doi.org/10.1093/bioinformatics/bty654>.
- 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>.
- 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. 2018. Porechop. <https://github.com/rwrick/Porechop>.
- Wick RR. 2018. FilTlong. <https://github.com/rwrick/FilTlong>.
- 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>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>.
- 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>.