



Draft Genome Sequence of *Bacillus* sp. strain YC2, an Isolate from the South African Medicinal Plant *Pelargonium sidoides*

Aza Z. Mqulwa,^a Wai Yin Chan,^{b,c} Lauren A. du Toit,^a  Kim M. Trollope,^a  Heinrich Volschenk^a

^aDepartment of Microbiology, Stellenbosch University, Stellenbosch, South Africa

^bDepartment of Biochemistry, Genetics and Microbiology (BGM) and Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

^cNational Institute for Communicable Disease, National Health Laboratory Service, Johannesburg, South Africa

ABSTRACT A rhizosphere-associated *Bacillus* species was isolated from *Pelargonium sidoides* DC (*Geraniaceae*) tubers, whose commercial extracts are used in respiratory tract infection treatment. Genomic data for *Bacillus* isolates associated with *Pelargonium sidoides* is lacking. Here, we report the draft genome sequence of *Bacillus* sp. strain YC2.

Pelargonium sidoides tuber extracts are used globally as commercial herbal medicines for respiratory infections (1). *In vitro* evidence detailed the antifungal activity of *P. sidoides* extracts against *Cryptococcus neoformans* (2), whose route of infection is via the inhalation of infectious particles (3, 4). Rhizospheric and endophytic microbes produce antimicrobial compounds and supply them to plant hosts or the surrounding areas. Plant growth-promoting *Bacillus* spp. are widely reported for their beneficial role in agriculture and their abundance in soil, plant rhizospheres, and within plant hosts (5). To expand the genomic information of rhizospheric and endophytic microbes, genome sequencing was performed on the *Bacillus* sp. strain YC2 associated with *P. sidoides* tubers.

Strain YC2 was isolated from serially diluted macerated suspensions of surface-sterilized, wild *P. sidoides* tubers (supplied by Parceval Pharmaceuticals, Wellington, South Africa) plated on lysogeny broth (LB) agar with 10 μ g/mL nalidixic acid and cycloheximide. The pure isolate was cultured from a single-colony 5-mL preculture in 50 mL LB at 37°C and 120 rpm. After 24 h, cells were harvested (4,863 $\times$ g, 20 min, 25°C), and genomic DNA was extracted (6). Briefly, cells were washed with 10 mL lysis buffer (0.5 M EDTA, 0.1 M NaCl, pH 7.5), resuspended in 4 mL lysis buffer, and treated with 10 mg lysozyme at 37°C for 10 min. Thereafter, 0.3 mL 20% (wt/vol) *N*-laurylsarcosine was added. After 5 min at 37°C, 4 mL phenol was added, followed by centrifugation at 2,330 $\times$ g for 20 min and 25°C. Supernatant was decanted and treated with 10 μ g RNase A for 2 h at 37°C. Genomic DNA was re-extracted with chloroform, ethanol precipitated, air-dried, and dissolved in 0.05 M Tris-EDTA buffer, pH 8. Concentration was determined by Qubit DNA analysis (Invitrogen, USA).

Genome sequencing was conducted by the Microbial Genome Sequencing Center in Pittsburgh, USA (<https://www.migscenter.com/>). Multiplexed Illumina sequencing libraries were constructed (7) and sequenced on the NextSeq 550 platform with high-output flowcell using the recommended Illumina manufacturer protocol. A total of 5,865,370 paired-end reads with a read length of 149 bp were generated. *De novo* assembly was conducted using SPAdes v3.12.0 (8) with the read coverage cutoff value of 80 $\times$ after the raw sequence data were trimmed by Trim Galore v0.6.6 (9, 10) (access date 9 February 2021). Average nucleotide identity based on BLAST (ANIb) with alignment fraction (AF%) was calculated using pyani v0.2.11 (11). The assembly was annotated by NCBI Prokaryotics Genome Annotation Pipeline v5.2 (12). Default parameters were used for all software unless otherwise noted.

Editor Leighton Pritchard, SIPBS, University of Strathclyde

Copyright © 2022 Mqulwa 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 Heinrich Volschenk, volschenk@sun.ac.za.

The authors declare no conflict of interest.

Received 8 September 2021

Accepted 22 February 2022

Published 28 March 2022

The draft genome of strain YC2 is 3,912,576 bp in size, with 44 contigs, N_{50} value of 275,976 bp, and a GC content of 45.23%. The assembly has a coverage of 334× and carries 3,876 protein coding sequences and 54 tRNA, 5 noncoding RNA (ncRNA), and 3 rRNA genes. The 16S rRNA sequence shares an 100% identity to *Bacillus amyloliquefaciens* strain AD2. The genome sequence shows ANIb values of 86.46 (82.5%) and 94.5 (84.4%) with *Bacillus amyloliquefaciens* ATCC 23350^T and *Bacillus nakamurai* CCUG68786^T, respectively. It contains 16 putative biosynthetic gene clusters for secondary metabolites as predicted by the antiSMASH v6.0.1 database (13) (access date, 31 August 2021).

Data availability. The draft genome assembly for *Bacillus* sp. YC2 has been deposited at the National Center for Biotechnology Information (NCBI) with GenBank accession number [JAHXDO000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JAHXDO000000000). The SRA, BioProject, and BioSample accession numbers are [SRR15685468](https://www.ncbi.nlm.nih.gov/sra/SRR15685468), [PRJNA748011](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA748011), and [SAMN20305144](https://www.ncbi.nlm.nih.gov/biosample/SAMN20305144), respectively.

ACKNOWLEDGMENTS

K.M.T. and A.Z.M. were supported by the National Research Foundation of South Africa (grant numbers 88180 and 126676, respectively).

REFERENCES

1. Makunga NP, Philander LE, Smith M. 2008. Current perspectives on an emerging formal natural products sector in South Africa. *J Ethnopharmacol* 119:365–375. <https://doi.org/10.1016/j.jep.2008.07.020>.
2. Samie S, Trollope KM, Joubert L-M, Makunga NP, Volschenk H. 2019. The antifungal and *Cryptococcus neoformans* virulence attenuating activity of *Pelargonium sidoides* extracts. *J Ethnopharmacol* 235:122–132. <https://doi.org/10.1016/j.jep.2019.02.008>.
3. Diamond RD. 1990. *Cryptococcus neoformans*, p 1980–1989. In Mandell GL, Douglas RG, Bennett JE (ed), *Principles and practice of infectious diseases*, 3rd ed. Churchill Livingstone, New York, NY.
4. Murphy JW. 1989. *Cryptococcosis*, p 93–138. In Cox RA (ed), *Immunology of fungal diseases*, 1st ed. CRC Press, Boca Raton, FL.
5. Ahmed EA, Hassan EA, El Tobgy KMK, Ramadan EM. 2014. Evaluation of rhizobacteria of some medicinal plants for plant growth promotion and biological control. *Ann Agric Sci* 59:273–280. <https://doi.org/10.1016/j.aos.2014.11.016>.
6. Harwood CR, Cutting SM. 1990. *Molecular biological methods for Bacillus*. John Wiley & Sons Ltd., Chichester, UK.
7. Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010:pdbs.ppt5448. <https://doi.org/10.1101/pdb.ppt5448>.
8. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics* 70:e102. <https://doi.org/10.1002/cpbi.102>.
9. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10. <https://doi.org/10.14806/ej.17.1.200>.
10. Krueger F, Galore T. 2015. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
11. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 8:12–24. <https://doi.org/10.1039/C5AY02550H>.
12. 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>.
13. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 39:W339–W346. <https://doi.org/10.1093/nar/gkr466>.