



Draft Genome Sequence of the Putative Endophytic Bacterium *Pantoea agglomerans* R6, Associated with *Lactuca serriola* from South Africa

Brandon Istain,^a Prudent S. Mokgokong,^c Ruomou Wu,^a Junaid Mia,^a Arun Gokul,^a David Mendoza-Cozatl,^{d,e} Morné Du Plessis,^c  Marshall Keyster^{a,b}

^aEnvironmental Biotechnology Laboratory, Department of Biotechnology, University of the Western Cape, Bellville, South Africa

^bDST-NRF Centre of Excellence in Food Security, University of the Western Cape, Bellville, South Africa

^cZoological Research, Foundational Research and Services, South African National Biodiversity Institute, Pretoria, South Africa

^dDepartment of Biotechnology, University of the Western Cape, Cape Town, South Africa

^eDivision of Plant Sciences, C.S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri, USA

ABSTRACT Here, we present the draft genome sequence (~4.7 Mb) of the endophytic bacterium *Pantoea agglomerans* strain R6, which was isolated from surface-sterilized roots of *Lactuca serriola* (prickly lettuce).

Pantoea agglomerans is a Gram-negative aerobic bacillus that belongs to the family *Erwiniaceae* and can be isolated from a wide variety of environments (1, 2). In plants, various strains were found to be either pathogens or commensals (3) and possess many beneficial agricultural traits (4).

In this study, *P. agglomerans* strain R6 was isolated from surface-sterilized roots of *Lactuca serriola*. The *L. serriola* root tissues were collected from the University of the Western Cape in Bellville, South Africa (33°56'06.0"S, 18°37'41.0"E), in April 2019. Clean root samples were sterilized with 70% ethanol for 3 min, 2.5% sodium perchlorate for 5 min, and 70% ethanol for 1 min and washed 5 times with autoclaved distilled water. The sterilized roots were crushed using a sterile mortar and pestle in a 0.9% saline (NaCl) solution, followed by incubation at 22°C for 5 h. The solution was diluted and spread plated (50 µl) onto R-2A agar (Sigma; catalog number 17209) and incubated at 22°C for 5 days. Following incubation, single colonies were picked and subcultured until a pure culture was obtained. For genomic DNA (gDNA) extraction, a single pure colony was cultured in nutrient broth (5 ml) at 22°C. The total genomic DNA was extracted with the GenElute bacterial genomic DNA kit (Sigma; catalog number NA2110). The purity and concentration of the DNA were measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

Genomic DNA (100 ng) was fragmented using the Covaris ultrasonicator. The DNA was end repaired using the Ion Xpress Plus fragment library kit (Thermo Fisher Scientific), and the 2% agarose E-gel system was used to size select fragments larger than 300 bp. The gDNA size was confirmed using the high-sensitivity D1000 DNA ScreenTape kit on the Agilent TapeStation 4200 (Agilent Technologies). Libraries were constructed using the Ion Xpress kit (Thermo Fisher Scientific) and quantified using the QuantStudio 12K system using the Ion library TaqMan quantitation kit (Thermo Fisher Scientific). The gDNA libraries were diluted to 100 pM and subjected to emulsion PCR using the Ion 520 and Ion 530 Kit-Chef (Thermo Fisher Scientific). The final enriched libraries were loaded onto the Ion 530 chip kit and subjected to sequencing on the Ion Torrent 5S GeneStudio system.

This generated 2,219,537 single-end reads with a mean read length of 280 bp. Default parameters were used for all software, unless otherwise stated. FastQC v0.10.1

Citation Istain B, Mokgokong PS, Wu R, Mia J, Gokul A, Mendoza-Cozatl D, Du Plessis M, Keyster M. 2021. Draft genome sequence of the putative endophytic bacterium *Pantoea agglomerans* R6, associated with *Lactuca serriola* from South Africa. *Microbiol Resour Announc* 10:e00023-21. <https://doi.org/10.1128/MRA.00023-21>.

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine

Copyright © 2021 Istain 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 Marshall Keyster, mkeyster@uwc.ac.za.

Received 11 January 2021

Accepted 18 February 2021

Published 11 March 2021

(5) was used for read quality assessment, where an average base quality of 30 was observed across all bases. The raw reads were trimmed for removal of 20 bp at the 5' end along with selection for reads between 220 bp and 300 bp, using Trimmomatic v0.36 (6). In addition, reads with low GC content (<45%) were also removed from further analysis to prevent the generation of spurious results.

Assembly was performed using SPAdes v3.13.0 (7). The final draft assembly contained 366 contigs with selection for only contigs greater than 1,000 bp. The assembly covered a total of 439, 942, and 830 bp with an N_{50} value of 25,969 and GC content of 55.61%. The depth of sequencing for the assembly was calculated at approximately 100× coverage. The 16S rRNA gene of R6 was used to perform a BLAST search against the NCBI nonredundant nucleotide database, which returned *Pantoea agglomerans* as the closest identity.

The genome sequence and associated annotation contribute to the expanding database of *Pantoea* species that have been identified to date (8). The output serves as the foundational data to further explore the utility of the features of the sequenced genome, particularly in the context of its endophytic relationship.

Data availability. The assembled genome sequence of *Pantoea agglomerans* R6 is available at GenBank under the accession number [JACRVA000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JACRVA000000000) (version 1) with BioProject number [PRJNA647595](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA647595), BioSample number [SAMN15595211](https://www.ncbi.nlm.nih.gov/biosample/SAMN15595211), and SRA accession number [SRR12904959](https://www.ncbi.nlm.nih.gov/sra/SRR12904959).

ACKNOWLEDGMENTS

This study was supported by the South African National Biodiversity Institute, the University of the Western Cape, and the University of Missouri-Columbia. Funding was provided by the National Research Foundation (NRF), DST-NRF Centre of Excellence in Food Security and Grain SA.

REFERENCES

1. Wright SA, Zumoff CH, Schneider L, Beer SV. 2001. *Pantoea agglomerans* strain EH318 produces two antibiotics that inhibit *Erwinia amylovora* in vitro. *Appl Environ Microbiol* 67:284–292. <https://doi.org/10.1128/AEM.67.1.284-292.2001>.
2. Cruz AT, Cazacu AC, Allen CH. 2007. *Pantoea agglomerans*, a plant pathogen causing human disease. *J Clin Microbiol* 45:1989–1992. <https://doi.org/10.1128/JCM.00632-07>.
3. Moloto VM, Goszczynska T, Hassen AI, Pierneef R, Coutinho T. 2020. Draft genome sequences of *Pantoea agglomerans* strains BD1274 and BD1212, isolated from onion seeds, reveal major differences in pathogenicity and functional genes. *Microbiol Resour Announc* 9:e01507-19. <https://doi.org/10.1128/MRA.01507-19>.
4. Dutkiewicz J, Mackiewicz B, Lemieszek MK, Golec M, Milanowski J. 2015. *Pantoea agglomerans*: a mysterious bacterium of evil and good. Part I. Deleterious effects: dust-borne endotoxins and allergens—focus on cotton dust. *Ann Agric Environ Med* 22:576–588. <https://doi.org/10.5604/12321966.1185757>.
5. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
6. 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>.
7. Anton B, Sergey N, Dmitry A, Alexey AG, Mikhail D, Alexander SK, Valery ML, Sergey IN, Son P, Andrey DP, Alexey VP, Alexander VS, Nikolay V, Glenn T, Max AA, Pavel AP. 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>.
8. Tambong JT. 2019. Taxogenomics and systematics of the genus *Pantoea*. *Front Microbiol* 10:2463. <https://doi.org/10.3389/fmicb.2019.02463>.