

# Complete genome sequence of Australian soil bacterium *Rouxiella badensis* DAR84756 resolved with Oxford Nanopore long-read and Illumina sequences

Saina Paul,<sup>1</sup> Peter J. Anderson,<sup>1</sup> Gregg J. Maynard,<sup>1</sup> Michael Dyall-Smith,<sup>2</sup> Timothy Kudinha<sup>3</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** The complete genome sequence of the bacterium *Rouxiella badensis* DAR84756, isolated from soil in Orange, NSW, Australia, was resolved using a combination of Nanopore long-read and Illumina short-read sequencing. The genome consists of a single, circular chromosome of 5,004,491 bp and a plasmid of 40,722 bp.

**KEYWORDS** soil, rhizosphere, Gram negative

*Rouxiella badensis* is a Gram-negative bacterium in the family *Yersiniaceae* (1) found associated with soil and plant ecosystems (2), with emerging applications including post-harvest preservation of berries (3), a human antidepressant probiotic (4–6), biosynthesis of L-asparaginase (7), and uncharacterized antibiotics (8). High-quality genome sequences, such as described here, allow improved identification of metabolic pathways and secondary metabolite prediction (9, 10).

This study resolves the genome of *R. badensis* DAR84756, isolated in 2011 from a soil sample from Orange, NSW, Australia (DD latitude –33.2452, longitude 149.11586). Isolation was by addition of equal volume sterile ultrapure water to soil, which was diluted 1:10, plated onto potato dextrose agar (PDA), and incubated at 25°C for 24 hours. Single colonies were isolated by streak plate onto PDA, from which reference –80°C freezer stocks were produced and stored in 50% glycerol/PD broth.

DNA for Illumina and Nanopore libraries was prepared by the same method, from separate genomic DNA extractions, each from a single colony inoculated into 2 mL of PD broth, incubated overnight at 28°C in a shaking incubator at 200 rpm. Genomic DNA preparations were made from this with the Wizard genomic DNA purification kit (Promega, Australia) following the manufacturer's instructions for Gram-negative bacteria. DNA quality and quantity were assessed by UV spectrometry and agarose gel electrophoresis. Frozen DNA was sent for library preparation and sequencing.

The Nanopore library was generated using the SQK-LSK109 kit (Oxford Nanopore, UK) without DNA shearing and sequenced on the GridION platform using a FLO-MIN106 (R9.4.1) flow cell at the Garvan Institute of Medical Research, Australia. Reads were called and quality controlled by MinKNOW version 20.06.17 and Guppy version 4.0.11 (11) with default settings. The Illumina DNA library was generated by fractionation and size selection on Pippin Prep (Sage Science, USA) with sequence generated on the MiSeq platform producing 2,885,611 paired reads, each 251 bp in length.

Genome assembly was completed in two sequential steps. The first was *de novo* assembly of the genome using Flye assembler (12) version 1.2 with Nanopore long DNA sequences, running on Geneious Prime 2022.1.1 software platform with default settings, except polishing was set to five iterations. Flye assembler resolved two circular contigs: a chromosome of 5,002,930 bp and a plasmid of 40,696 bp.

**Editor** J. Cameron Thrash, University of Southern California, Los Angeles, California, USA

Address correspondence to Peter J. Anderson, ptanderson@csu.edu.au.

The authors declare no conflict of interest.

**Received** 25 September 2023

**Accepted** 9 November 2023

**Published** 1 December 2023

Copyright © 2023 Paul 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/).

TABLE 1 Summary of DNA sequence assembly data

Description	<i>R. badensis</i>	Plasmid
	DAR84756	Sp1
GenBank accession no.	<a href="#">CP114062.1</a>	<a href="#">CP114061.2</a>
Chromosome size (bp)	5,004,491	40,722
GC content (%)	52.80%	49.30%
ONT sequencing		
No. of reads	268,783	
Read length (mean) (bp)	7,868	
Read $N_{50}$	19,429	
Illumina sequencing		
No. of reads	2,885,611	
Read length (bp)	2 × 251	
Genome statistics		
Mean coverage (x)	414	
tRNAs	78	
rRNA (5S, 16S, 23S)	8, 7, 7	
Pseudo genes	39	
Genes (RNA)	110	
Coding genes	4,466	
Total genes	4,615	

In the second step, Illumina 251-bp shotgun DNA reads were aligned against the Flye assembly using the “Map to reference” option in Geneious (13) with default parameters, allowing correction of Nanopore sequencing misreads, calls, and indels (14) by leveraging the accuracy of the Illumina short reads, resulting in 1,610 changes across the entire genome. The final, polished genome consisted of a circular chromosome of 5,004,491 bp and a plasmid of 40,722 bp (see Table 1).

*R. badensis* DAR84756 whole genome average nucleotide identity was 99.63% compared to *R. badensis* DSM100043<sup>T</sup> (accession number [CP114060.1](#)) confirming it to be the same species (15, 16). However, plasmid sizes differed significantly being 40,722 bp and 116,582 bp, respectively.

The NCBI Prokaryotic Genome Annotation Pipeline (17) v6.4 annotation predictions and summary are shown in Table 1.

## ACKNOWLEDGMENTS

This research was supported by a New South Wales Health and Charles Sturt University PhD scholarship to Saina Paul.

## AUTHOR AFFILIATIONS

<sup>1</sup>School of Dentistry and Medical Sciences, Charles Sturt University, Leeds Parade, Orange, New South Wales, Australia

<sup>2</sup>Veterinary Biosciences, Melbourne Veterinary School, Faculty of Science, University of Melbourne, Parkville, Victoria, Australia

<sup>3</sup>NSW Health Pathology West, Orange, New South Wales, Australia

## AUTHOR ORCIDs

Peter J. Anderson  <http://orcid.org/0000-0001-6577-7142>

Michael Dyal-Smith  <http://orcid.org/0000-0002-1880-1960>

## AUTHOR CONTRIBUTIONS

Saina Paul, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing | Peter J. Anderson,

Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing | Gregg J. Maynard, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review and editing | Michael Dyall-Smith, Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Timothy Kudinha, Conceptualization, Methodology, Project administration, Supervision, Writing – review and editing

## DATA AVAILABILITY

The whole genome sequence of *R. badensis* DAR84756 has been deposited in GenBank under the accession no. [CP114062.1](https://doi.org/10.1093/ncbi/CP114062.1) and for its plasmid Sp1 [CP114061.2](https://doi.org/10.1093/ncbi/CP114061.2) (18) with the BioSample accession number of [SAMN32064793](https://www.ncbi.nlm.nih.gov/biosample/SAMN32064793). The associated BioProject number for this project is [PRJNA909035](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA909035). The raw Illumina data for this genome are available from NCBI with accession number [SRX19145821](https://www.ncbi.nlm.nih.gov/sra/SRX19145821). The raw Oxford Nanopore data are available through accession number [SRX19145820](https://www.ncbi.nlm.nih.gov/sra/SRX19145820).

## REFERENCES

- Adeolu M, Alnajjar S, Naushad S, Gupta R. 2016. "Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacteriales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov". *Int J Syst Evol Microbiol* 66:5575–5599. <https://doi.org/10.1099/ijsem.0.001485>
- Yahfoufi N, Alsadi N, Mallet JF, Kulshreshtha G, Hincke M, Ismail N, Matar C. 2021. Immunomodulation and intestinal morpho-functional aspects of a novel gram-negative bacterium *Rouxiella badensis* subsp. *acadiensis*. *Front Microbiol* 12:569119. <https://doi.org/10.3389/fmicb.2021.569119>
- Morales-Cedeño LR, de Los Santos-Villalobos S, Santoyo G. 2021. Functional and genomic analysis of *Rouxiella badensis* SER3 as a novel biocontrol agent of fungal pathogens. *Front Microbiol* 12:709855. <https://doi.org/10.3389/fmicb.2021.709855>
- Novotny-Nuñez I, Perdigón G, Matar C, Martínez Monteros MJ, Yahfoufi N, Cazorla SI, Maldonado-Galdeano C. 2023. Evaluation of *Rouxiella badensis* subsp. *acadiensis* (Canan SV-53) as a potential probiotic bacterium. *Microorganisms* 11:1347. <https://doi.org/10.3390/microorganisms11051347>
- Yahfoufi N, Ah-Yen EG, Chandrasegaram R, Aly S, Murack M, Kadamani AK, Matar C, Ismail N. 2021. Adolescent use of potential novel probiotic *Rouxiella badensis* subsp. *acadiensis* (Canan SV-53) mitigates pubertal LPS-induced behavioral changes in adulthood in a sex-specific manner by modulating 5HT1A receptors expression in specific brain areas. *Compr Psychoneuroendocrinol* 7:100063. <https://doi.org/10.1016/j.cpnec.2021.100063>
- Yahfoufi N, Kadamani AK, Aly S, Al Sharani S, Liang J, Butcher J, Stintzi A, Matar C, Ismail N. 2023. Pubertal consumption of *R. badensis* subspecies *acadiensis* modulates LPS-induced immune responses and gut microbiome dysbiosis in a sex-specific manner. *Brain Behav Immun* 107:62–75. <https://doi.org/10.1016/j.bbi.2022.09.013>
- Gilavand F, Kavyanifard A, Marzban A. 2018. L-asparaginase-producing *Rouxiella* species isolation, antileukemia activity evaluation, and enzyme production optimization. *Res Mol Med* 6:29–44.
- Nam YH, Hwang BS, Choi A, Chung EJ. 2020. Isolation and characterization of strain *Rouxiella* sp. S1S-2 producing antibacterial compound. *Korean J Microbiol* 56:152–159.
- Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, Escobar J, Flowers D, Fotopulos D, Garcia C, Gomez M, Gonzales E, Haydu L, Lopez F, Ramirez L, Retterer J, Rodriguez A, Rogers S, Salazar A, Tsai M, Myers RM. 2004. Quality assessment of the human genome sequence. *Nature* 429:365–368. <https://doi.org/10.1038/nature02390>
- Salzberg SL, Yorke JA. 2005. Beware of mis-assembled genomes. *Bioinformatics* 21:4320–4321. <https://doi.org/10.1093/bioinformatics/bti769>
- 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>
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Lu H, Giordano F, Ning Z. 2016. Oxford Nanopore Minion sequencing and genome assembly. *Genom Proteom Bioinform* 14:265–279. <https://doi.org/10.1016/j.gpb.2016.05.004>
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91. <https://doi.org/10.1099/ijms.0.64483-0>
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
- Tatusova T, DiCuccio M, Badretdin A, Chetverin 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>
- Paul S, Anderson PJ, Maynard G, Dyall-Smith M, Kudinha T. 2022. Complete genome sequence of an Australian strain of *Rouxiella badensis* DAR84756 and resolution of the *R. badensis* DSM100043 and *R. chamberiensis* DSM28324 genomes. Available from: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA909035>