

# High-quality draft genome sequence of *Paenibacillus* sp. RC80, a candidate for biofuel production

Gabriella Rizzo,<sup>1</sup> Mallory Choudoir,<sup>1,2</sup> Rachel Simoes,<sup>1</sup> Nipuni Dayarathne,<sup>1</sup> Kristen M. DeAngelis<sup>1</sup>

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

**ABSTRACT** *Paenibacillus* sp. RC80 was isolated from temperate deciduous forest soil in New England. The assembled genome is a single contig with 5,977,337 bp and 97.15% estimated completion. RC80 contains features for 2,3-butanediol dehydrogenase production and pathways involved in ethanol production.

**KEYWORDS** soil microbiology, environmental microbiology, bioinformatics

The identification of RC80 as a potentially novel species within the *Paenibacillus* genus hints at the vast reservoir of untapped soil microbial diversity, offering tantalizing prospects for biotechnological applications, particularly in the realm of biofuel production (1). Therefore, we sequenced a bacterial genome isolated from temperate forest soil (2) to understand the genetic profile of this novel species and its potential for biofuel production.

RC80 was isolated in 2022 from mineral soils collected, air-dried, and archived in 1998 from the Harvard Forest (2), a temperate, deciduous forest in Petersham, MA (42.54°N, 272.18°W). RC80 was isolated on ISP2 medium (3) and later grown on 10% tryptic soy agar (TSA), both at 25°C. For gDNA extraction, a single colony was grown in 10% tryptic soy broth (TSB) liquid media at 30°C, shaking at 150 rpm until an optical density of 0.5 was reached. Cells were pelleted using centrifugation at 4,000 rpm for 15 minutes, and genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (4). The genome was sequenced on MinION (Oxford Nanopore Technology) using Nanopore R9 flow cells (R9.4.1) with SQK-LSK109 sequencing kit by SeqCenter (Pittsburgh, PA). High-accuracy base calling with Guppy v.4.5.4 was used to achieve Q20 performance. The DNA was not sheared or size-selected. The 16S rRNA gene was sequenced (4), and BLASTn (default settings) search (5) against the National Center for Biotechnology Information nucleotide database determined the closest cultured representative of RC80 to be *Paenibacillus polymyxa* SC2 (98.8% identity).

The genome was assembled, annotated, and analyzed as part of the Bioinformatics Lab (MICROBIO 590B) course at University of Massachusetts Amherst (6). All software was deployed using default parameters unless otherwise specified. The raw read N50 for the Nanopore sequencing was 8,054 bp as determined by SeqFu (7). This generated 1,463,430,378 bp in 288,192 reads. The reads were subsampled to 40× target coverage (240,000,000 bp) and quality controlled using Filtlong v.0.2.1 (8). A *de novo* assembly was generated using Flye v.2.9.1 (9) followed by a consensus sequence using Minimap2 v.2.24 (10) and Racon v.1.4.3 (11). The final polishing was performed using Medaka v.1.7.2 (12). The genome was not trimmed, rotated, or circularized. Quality assessment of the sequence was evaluated using both QUAST v.4.4 (13) and CheckM v.1.0.18 (14). The annotation of the genome was completed using RAST v.1.073 (15) and Prokka v.1.14.5 (16) on Kbase (17). The final draft assembly contained one contig with an N50 of 5,977,337 bp, was estimated 97.15% complete and 0.07% contaminated, and contained nine copies each of 23S, 16S, and 5S ribosomal RNA genes and 71 tRNA copies. Therefore,

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

Address correspondence to Kristen M. DeAngelis, deangelis@umass.edu.

The authors declare no conflict of interest.

See the funding table on p. 2.

**Received** 30 November 2023

**Accepted** 10 May 2024

**Published** 7 June 2024

Copyright © 2024 Rizzo 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/).

the genome assembly is considered high quality (18). The average nucleotide identity (ANI) estimate (19) of *Paenibacillus polymyxa* SC2 was 85.43%. Since the ANI number is less than 95%, it is likely that this genome is a novel cultured species of *Paenibacillus*.

The RAST annotated genome shows that RC80 has 2,3-butanediol dehydrogenase alcohol forming feature. Additionally, RC80 is likely capable of producing ethanol because it has enzymes such as alcohol dehydrogenase present in the glycolysis pathway, determined using the Build Metabolic Model app v.2.0.0 (20). When produced in combination, 2,3-butanediol and ethanol make RC80 a potential candidate for biotechnological purposes, such as the production of biofuels (21).

## ACKNOWLEDGMENTS

This project was conducted with support by National Science Foundation (NSF), grant number: DEB-1749206. The isolate was collected as part of the microbiology of climate change study in the Harvard Forest long-term soil warming experiment, maintained with support from the NSF Long Term Ecological Research Program, grant number: DEB-1832110 and Long Term Research in Environmental Biology, grant numbers: DEB-1456610 and DEB-1949882. We are grateful to Dr. Serita Frey and Dr. Mel Knorr for maintaining this infrastructure for our community.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Microbiology, University of Massachusetts, Amherst, Massachusetts, USA

<sup>2</sup>Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

## AUTHOR ORCIDs

Gabriella Rizzo  <http://orcid.org/0009-0007-8415-9308>

Mallory Choudoir  <http://orcid.org/0000-0002-9117-5150>

Kristen M. DeAngelis  <http://orcid.org/0000-0002-5585-4551>

## FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	DEB-1749206, DEB-1832110, DEB-1456610	Gabriella Rizzo Mallory Choudoir Rachel Simoes Nipuni Dayarathne Kristen M. DeAngelis

## DATA AVAILABILITY

The static narrative for RC80 is available on [KBase](#). The [RAST](#) and [Prokka](#) annotations are available on FigShare. The 16S rRNA gene sequence accession number for RC80 is [OR690258.1](#). The raw whole genome sequence reads are available in GenBank under the BioProject accession number [PRJNA1029665](#). The BioSample accession number for RC80 is [SAMN38017374](#). The Sequence Read Archive (SRA) accession number for RC80 is [SRR26593010](#). The complete genome reference number for RC80 is [CP137860](#).

## REFERENCES

- Grady EN, MacDonald J, Liu L, Richman A, Yuan Z-C. 2016. Current knowledge and perspectives of *Paenibacillus*: a review. *Microb Cell Fact* 15:203. <https://doi.org/10.1186/s12934-016-0603-7>
- Melillo JM, Frey SD, DeAngelis KM, Werner WJ, Bernard MJ, Bowles FP, Pold G, Knorr MA, Grandy AS. 2017. Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science* 358:101–105. <https://doi.org/10.1126/science.aan2874>
- Shirling EB, Gottlieb D. 1966. Methods for characterization of streptomyces species1. *Int J Syst Evol Microbiol* 16:313–340. <https://doi.org/10.1099/00207713-16-3-313>
- DeAngelis KM, Pold G, Topçuoğlu BD, van Diepen LTA, Varney RM, Blanchard JL, Melillo J, Frey SD. 2015. Long-term forest soil warming alters microbial communities in temperate forest soils. *Front Microbiol* 6:104. <https://doi.org/10.3389/fmicb.2015.00104>

5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
6. DeAngelis K, Choudoir M, Eng A, Morrow M. 2023. Bioinformatics lab: a course-based undergraduate research experience curriculum. *Microbiol Educ Mater*. <https://doi.org/10.7275/g0fg-5939>
7. Telatin A. 2020. SeqFu docs · SeqFu - FASTX sequence utilities
8. Wick R. 2023. rrwick/Filtlong. C++
9. 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>
10. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
11. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>
12. Medaka. 2023. Python. Oxford Nanopore Technologies.
13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
14. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
15. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>
16. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
17. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>
18. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloë-Fadrosh EA, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>
19. 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>
20. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28:977–982. <https://doi.org/10.1038/nbt.1672>
21. Adlakha N, Pfau T, Ebenhöf O, Yazdani SS. 2015. Insight into metabolic pathways of the potential biofuel producer, *Paenibacillus polymyxa* ICGEB2008. *Biotechnol Biofuels* 8:159. <https://doi.org/10.1186/s13068-015-0338-4>