

High-quality genomes of *Paenibacillus* spp. RC334 and RC343, isolated from a long-term forest soil warming experiment

Claire E. Kitzmiller,¹ Wyatt C. Tran,¹ Brendan Sullivan,¹ Florencia Cortez,¹ Mallory Choudoir,^{1,2} Rachel Simoes,¹ Nipuni Dayarathne,¹ Kristen M. DeAngelis¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Paenibacillus* spp. RC334 and RC343 were isolated from heated soil in a long-term soil warming experiment. Both genomes were 5.98 Mb and assembled as a single contig. We describe the assembly and annotation of the two high-quality draft genomes for these isolates here.

KEYWORDS soil microbiology, climate change

Soil microbes have the potential to mitigate the impact of global warming, but we don't understand how they respond to long-term warming (1, 2). We sequenced bacterial genomes isolated from the Harvard Forest soil warming experiment (2) to understand the genetic profile of these novel species.

Paenibacillus spp. RC334 and RC343 were isolated in 2022 on Actinobacteria Isolation Agar (3) with 100 mg L-1 cycloheximide from mineral soil from a heated soil plot (43.54N 72.18 W), collected 10 cm below the surface, at an elevation of 355 m, using a steel corer. The isolates were streaked, and single colonies were grown in 10% tryptone soy broth media at 30°C shaking at 150 rpm until OD of 0.5. gDNA was extracted by CTAB (4) for RC343 and using the Blood & Tissue DNEasy Kit (Qiagen) following the manufacturer's instructions for RC334. Libraries were prepared with the Ligation Sequencing Kit SQK-LSK-109, and samples were multiplexed using the Native Barcoding Expansion Kit EXP-NBD104. Oxford Nanopore sequencing was performed at SeqCenter (Pittsburgh, PA) using R9 flow cells (R9.4.1). High accuracy base calling with Guppy v4.5.4 was used to achieve Q20 performance.

The genomes were assembled, annotated, and analyzed as part of the Bioinformatics Lab (MICROBIO 590B) course at the University of Massachusetts Amherst (5). FiltLong (6) was run to remove low-quality reads and specified a 40× coverage for RC334 and for RC343. The genomes were assembled *de novo* using Flye (7), and then Minimap2 (8) mapped the genome and completed pairwise alignment. Racon (9) created a genomic consensus, and Medaka (10) polished consensus sequences. Quast (11) and CheckM (12) were then used to assess the quality of the assembly. The genome assemblies for RC334 and RC343 are both of high quality (Table 1) (13).

The final assemblies were uploaded to KBase for analysis and annotation (14). All apps were run on the default settings unless otherwise indicated. Genomes were annotated using Prokka (Annotate Assembly and Re-annotate Genomes with Prokka—v1.14.5) (15) and classified using GTDB-Tk v1.7.0 (16) which assigns a taxonomic classification to the organism using domain-specific, concatenated proteins. Both genomes' domain is Bacteria, the phylum is Firmicutes, the class is Bacilli, the order is Paenibacillales, the family is Paenibacillaceae, the genus is *Paenibacillus*, and the species is *Paenibacillus terrae_A*.

The nearest neighbor for both genomes was identified to be *Paenibacillus polymyxa* SC2 using the phylogenetic tree made with a comparison of 49 clusters of Orthologous

Editor David A. Baltrus, The University of Arizona, Tucson, Arizona, USA

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

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 5 May 2023

Accepted 27 July 2023

Published 28 August 2023

Copyright © 2023 Kitzmiller 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 Genome assembly details from Quast

Features	RC334	RC343
Total base pairs in the assembly (bp)	183,313,012	244,003,914
Assembled genome size (bp)	5,979,552	5,982,416
Fold-coverage (total bp/genome size)	30.6	40.8
Assembly N50 (bp)	5,979,552	5,982,416
Assembly N75 (bp)	5,979,552	5,982,416
Number of contigs	1	1
G + C content (%)	46.72	46.71
Completion (%)	99.12	96.8
Contamination (%)	0.07	0.07

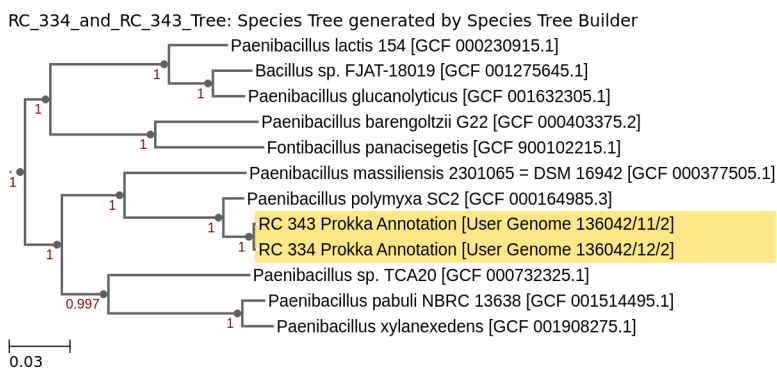


FIG 1 Phylogenetic tree based on 49 core genes. The phylogenetic tree was created on KBase (Insert Genome into SpeciesTree) (19) using a set of 49 core, universal genes defined by COG.

Groups (COG genes) (Fig. 1). Compute Average Nucleotide Identity (ANI) with FastANI v0.1.3 (17, 18) calculated the Average Nucleotide Identity between *P. polymyxa* SC2 and RC334 and RC343 to be 85.4% and 85.3%, respectively. RC334 and RC343 both had 94.2% ANI with a user *P. terrae* (GCF_000235585.1_assembly) genome assembly and annotation. When RC343 was compared to RC334, the ANI estimate was 99.95%. Since the ANI number is less than 95%, it is likely that these compared genomes are from different species than *P. polymyxa* and *P. terrae*. Further research into these isolates may provide deeper insight into the role of *Paenibacillus* sp. in microbial climate change responses.

ACKNOWLEDGMENTS

This project was conducted with support of a grant from the National Science Foundation (No. DEB-1749206). The soil warming experiments at Harvard Forest are maintained with support from the National Science Foundation Long Term Ecological Research Program (DEB-1832110) and a Long Term Research in Environmental Biology grant (DEB-1456610).

We are grateful to Drs. Serita Frey and Mel Knorr for maintaining this infrastructure for our community. Thank you to Ashely Eng for your help and support throughout.

AUTHOR AFFILIATIONS

¹Department of Microbiology, University of Massachusetts, Amherst, Massachusetts, USA

²Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

AUTHOR ORCIDs

Claire E. Kitzmiller  <http://orcid.org/0009-0001-7069-4891>

Wyatt C. Tran  <http://orcid.org/0009-0002-6162-6261>
 Brendan Sullivan  <http://orcid.org/0009-0005-6836-1552>
 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-1456528, DEB-1832210	Kristen M. DeAngelis

AUTHOR CONTRIBUTIONS

Claire E. Kitzmiller, Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft | Brendan Sullivan, Formal analysis, Investigation, Writing – review and editing | Florencia Cortez, Formal analysis | Mallory Choudoir, Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing | Rachel Simoes, Data curation | Nipuni Dayarathne, Data curation | Kristen M. DeAngelis, Conceptualization, Formal analysis, Funding acquisition, Resources, Supervision, Writing – review and editing.

DATA AVAILABILITY

The 16S rRNA gene sequence accession number for RC343 is [OQ547097](https://doi.org/10.1101/2024.01.05.574709). The 16S rRNA gene sequence accession number for RC334 is [OQ547098](https://doi.org/10.1101/2024.01.05.574708). The raw whole-genome sequence reads are available in GenBank under the BioProject accession number [PRJNA949990](https://doi.org/10.1101/2024.01.05.574999). The BioSample accession number for RC334 is [SAMN33990111](https://doi.org/10.1101/2024.01.05.574911) and for RC343 is [SAMN33990112](https://doi.org/10.1101/2024.01.05.574912). The Sequence Read Archive (SRA) accession number for RC334 is [SRR24019814](https://doi.org/10.1101/2024.01.05.574914) and for RC343 is [SRR24019813](https://doi.org/10.1101/2024.01.05.574913). The draft genome reference number for RC334 is [NZ_CP125370.1](https://doi.org/10.1101/2024.01.05.574970) and for RC343 is [CP125371.1](https://doi.org/10.1101/2024.01.05.574971).

REFERENCES

- Friedlingstein P, Cox P, Betts R, Bopp L, von Bloh W, Brovkin V, Cadule P, Doney S, Eby M, Fung I, Bala G, John J, Jones C, Joos F, Kato T, Kawamiya M, Knorr W, Lindsay K, Matthews HD, Raddatz T, Rayner P, Reick C, Roeckner E, Schnitzler K-G, Schnur R, Strassmann K, Weaver AJ, Yoshikawa C, Zeng N. 2006. Climate–carbon cycle feedback analysis: results from the C4MIP model Intercomparison. *J Clim* 19:3337–3353. <https://doi.org/10.1175/JCLI3800.1>
- 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>
- Lechevalier MP, Lechevalier HA. 1974. *Nocardia Amarae* SP. Nov., an Actinomycete common in foaming activated sludge. *Int j syst bacteriol* 24:278–288. <https://doi.org/10.1099/00207713-24-2-278>
- 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>
- Bioinformatics lab: a course-based undergraduate research experience C" by Kristen M. Deangelis, Mallory Choudoir et al. 2023. Available from: https://scholarworks.umass.edu/micro_ed_materials/3/
- Wick R. 2022. C++. [rrwick/Filterlong](https://github.com/rrwick/Filterlong)
- 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>
- Li H. 2022. Minimap2: Pairwise alignment for nucleotide sequences | Bioinformatics | Oxford academic. Available from: <https://academic.oup.com/bioinformatics/article/34/18/3094/4994778?login=false>
- 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>
2022. Medaka. Python. Oxford Nanopore Technologies.
- Mikheenko A, Valin G, Prijbelski A, Saveliev V, Gurevich A. 2016. Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics* 32:3321–3323. <https://doi.org/10.1093/bioinformatics/btw379>
- 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>
- Bowers RM, Kyrpidis NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloë-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooshep S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Etema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattai T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T, Genome Standards Consortium. 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>
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD,

- Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>
15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
 16. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH, Hancock J. 2020. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
 17. 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>
 18. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities | microbiology society. 2023. Available from: <https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijfs.0.64483-0#tab2>
 19. Price MN, Dehal PS, Arkin AP. 2010. Fasttree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>