

Article title

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

Authors

Claire E. Kitzmiller^a, Wyatt C. Tran^a, Brendan Sullivan^a, Florencia Cortez^a, Mallory Choudoir^{a,b}, Rachel Simoes^a, Nipuni Dayarathne^a, and Kristen M. DeAngelis^{a#}.

a. Department of Microbiology, University of Massachusetts, Amherst, Massachusetts, USA

b. Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA

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.

Running title

Paenibacillus spp. RC334 and RC343 genomes.

Corresponding author's email address

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

Announcement

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 10cm below surface, at an elevation of 355m, using a steel corer. The isolates were streaked and single colonies 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 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 University of Massachusetts Amherst (5). FiltLong (6) was run to remove low quality reads and specified a 40X coverage for RC334 and for RC343. The genomes were assembled *de novo* using Flye (7), 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 high quality (Table 1)(13).

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

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*.

Figure 1. Phylogenetic Tree Based on 49 Core Genes. The phylogenetic tree was created on KBase (Insert Genome into SpeciesTree)(17) using a set of 49 core, universal genes defined by COG (clusters of Orthologous Groups).

The nearest neighbor for both genomes was identified to be *Paenibacillus polymyxa* SC2 using the phylogenetic tree made with a comparison of 49 COG genes (Fig. 1). Compute ANI with FastANI v0.1.3 (18, 19) calculated the Average Nucleotide Identity between *Paenibacillus 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 *Paenibacillus 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.

Data Availability

The 16S rRNA gene sequence accession number for RC343 is [OQ547097](#). The 16S rRNA gene sequence accession number for RC334 is [OQ547098](#). The raw whole genome sequence reads are available in GenBank under the BioProject accession number [PRJNA949990](#). The BioSample accession number for RC334 is [SAMN33990111](#) and for RC343 is [SAMN33990112](#). The Sequence Read Archive (SRA) accession number for RC334 is [SRR24019814](#) and for RC343 is [SRR24019813](#). The draft genome reference number for RC334 is NZ_ [CP125370.1](#) and for RC343 is [CP125371.1](#).

Acknowledgements

This project was conducted with support by 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 help and support throughout.

References

1. Friedl, J., Cox P, Betts R, Bopp L, Bloh W von, 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,

93 Rayner P, Reick C, Roeckner E, Schnitzler K-G, Schnur R, Strassmann K, Weaver AJ, Yoshikawa C,
94 Zeng N. 2006. Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model
95 Intercomparison. J Clim 19:3337–3353.

96 2. Melillo JM, Frey SD, DeAngelis KM, Werner WJ, Bernard MJ, Bowles FP, Pold G, Knorr MA, Grandy
97 AS. 2017. Long-term pattern and magnitude of soil carbon feedback to the climate system in a
98 warming world. Science 358:101–105.

99 3. Olsen. Personal Communication.

100 4. DeAngelis KM, Pold G, Topçuoğlu BD, van Diepen LTA, Varney RM, Blanchard JL, Melillo J, Frey SD.
101 2015. Long-term forest soil warming alters microbial communities in temperate forest soils. Front
102 Microbiol 6.

103 5. “Bioinformatics Lab: A Course-Based Undergraduate Research Experience C” by Kristen M.
104 DeAngelis, Mallory Choudoir et al. https://scholarworks.umass.edu/micro_ed_materials/3/.
105 Retrieved 3 May 2023.

106 6. Wick R. 2022. rrwick/Filtlong. C++.

107 7. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat
108 graphs. 5. Nat Biotechnol 37:540–546.

109 8. Li H. Minimap2: pairwise alignment for nucleotide sequences | Bioinformatics | Oxford Academic.
110 <https://academic.oup.com/bioinformatics/article/34/18/3094/4994778?login=false>. Retrieved 28
111 November 2022.

112 9. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from
113 long uncorrected reads. Genome Res 27:737–746.

114 10. 2022. Medaka. Python. Oxford Nanopore Technologies.

115 11. Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for de novo
116 assembly evaluation. Bioinformatics 32:3321–3323.

- 117 12. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the
118 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*
119 25:1043–1055. Accessed 22 January 2023.
- 120 13. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J,
121 Rivers AR, Elie-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom
122 RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G,
123 Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis
124 KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R,
125 Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat
126 Eren A, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single
127 amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and
128 archaea. 8. *Nat Biotechnol* 35:725–731.
- 129 14. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F,
130 Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari
131 S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA,
132 Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison
133 BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM,
134 Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP,
135 Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B,
136 Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC,
137 Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S,
138 Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. 7.
139 *Nat Biotechnol* 36:566–569.
- 140 15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinforma Oxf Engl* 30:2068–

141 2069. Accessed 22 January 2023.

142 16. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes
143 with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. Accessed 22 January 2023.

144 17. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for
145 Large Alignments. *PLoS ONE* 5:e9490. Accessed 22 January 2023.

146 18. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI
147 analysis of 90K prokaryotic genomes reveals clear species boundaries. 1. *Nat Commun* 9:5114.
148 Accessed 24 January 2023.

149 19. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities |
150 Microbiology Society.
151 <https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijjs.0.64483-0#tab2>.
152 Retrieved 5 April 2023.