 doi:10.20944/preprints201910.0263.v1

Unraveling the draft genome sequence of *Paenibacillus* sp. tmac-D7 a non-rhizobial endophyte from *Trifolium macraei* located in Bodega Bay, California

Richard Allen White III,[1,5-6] Emily E. Mclachlan,[1] Joseph P. Dunham,[4] Aaron Garoutte,[3] Maren L. Friesen[1-3]

[1]Department of Plant Pathology, Washington State University, Pullman, WA USA

[2]Department of Crop and Soil Sciences, Washington State University, Pullman, WA USA

[3]Department of Plant Biology, Michigan State University, East Lansing, MI, USA

[4]SeqOnce LLC Pasadena, CA, USA

[5]RAW Molecular Systems (RMS) LLC, Spokane, WA USA

[6]Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW, Australia

Correspondence: Maren L. Friesen, Department of Plant Pathology, Crop and Soil Sciences, Washington State University. Email: m.friesen@wsu.edu

Abstract: *Paenibacillus* sp. tmac-D7 was isolated from coastline growing *Trifolium macraei* (double-head clover) root nodules from Bodega Bay, California. The draft genome is 5,567,337 bp with a G+C% of 52.4%, an N50 of 114,261 bp, and 5,282 predicted protein-coding genes. *Paenibacillus,* while found in many other environments, is frequently isolated from root nodules, with many acting as plant pathogen antagonists. *Paenibacillus* sp. tmac-D7 is the first genome of a non-rhizobial endophyte isolate from wild *Trifolium macraei* (double-head clover).

Data set: OSF repo https://osf.io/kcj9h/

Keywords: *Paenibacillus*, rhizosphere, non-rhizobial endophyte (NRE), *Trifolium*, clover, Bodega Bay, diazotroph, plant-growth promoting bacteria, free-living nitrogen fixer

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 October 2019 doi:10.20944/preprints201910.0263.v1

1. Summary

Root associated rhizosphere microbiome is a highly dynamic interface containing >30,000 bacterial species (**1-3**). Rhizobium are symbiotic nitrogen-fixing bacteria that form plant-accessory structures with leguminous plants (e.g., *Trifolium*) known as nodules, which biological nitrogen fixation the process of transforming molecular nitrogen (N_2) into ammonium (NH_3) occurs (**4**). While the association between rhizobia and plant forms nodule formation in legumes, the nodule is not a monoculture of a single rhizobium but contains a more extensive microbiome of non-rhizobial endophytes (**5**). *Paenibacillus,* while found in many other environments, is frequently isolated from root nodules (**5**). Many strains act as fungal plant pathogen antagonists (e.g., *Rhizoctonia* and *Fusarium*) as well as have antinematodal activity (**5-6**).

The 3,747,887 cleaned resulting paired-end reads were *de novo* assembled into 142 contigs, with a genome size of 5,567,337 bp with a G+C% of 52.4%, an N50 of 114,261 bp. Of the 142 contigs, 121 were >1 kbp, and 90 were >5 kbp. The resulting genome was 99.73% complete, with 1.88% contamination. The genome annotation predicts 87 tRNAs, 1 tmRNAs, 77 non-coding RNAs, 1 copy of 5S-16S-23S operons, 0 CRISPRs, and 5,282 predicted protein-coding genes.

Paenibacillus sp. tmac-D7 is the first genome isolate from wild *Trifolium macraei* (double-head clover) and is a member of the nodule non-rhizobial endosphere microbiome. Our genome elucidated here provides further genomic context to nodule microbiome, which could be used for rhizosphere engineering of food crops (e.g., soybean) or other agriculture-related plant growth-promoting applications (**7**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 October 2019 doi:10.20944/preprints201910.0263.v1

2. Data description

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession VIDW00000000. The version described in this paper is version VIDW00000000. Fourteen contigs were removed for being too small (< 200 bp) for GenBank submission. Raw data, contigs, and annotations for this genome can be found at https://osf.io/kcj9h/. The code used to generate assembly can be found at www.github.com/friesenlab/Paenibacillus_tmac-D7.

3. Methods

Paenibacillus sp. tmac-D7 was isolated from a *Trifolium macraei* (double-head clover) nodule from Bodega Bay, California. Nodule material was enriched on Tryptone-Agar (TA) plates after an incubated at 30° C for three days, a single colony was isolated (i.e., tmac-D7) then streaked for isolation on TA plates. MasterPure DNA extraction kit with proteinase K (Epicentre, Madison WI, USA) following manufacturers protocols. Illumina library was prepared using the SeqOne RhinoSeq kit following the manufacturer's protocols (https://seqonce.com/rhinoseq/). Michigan State University Research Technology Support Facility (RTSF) sequencing core completed DNA sequencing, library quantification, and sequenced on HiSeq 4000 150 bp paired-end read format.

Unless otherwise specified default parameters were used for all software with all versions of software listed. The Illumina sequencing data quality filtered and decontaminated using ATLAS (version 1.0) then assembled with using Unicycler (version 0.4.7) with default Illumina assembly parameters (**8-9**). CheckM (version 1.0.12) was used to estimate completeness and contamination (**10**). Annotation was completed using Prokka (version 1.13.3) with -rfam flag to obtain rRNAs and tmRNAs (**11**).

Acknowledgments

We acknowledge support from NSF IOS 1331218 and NSF DEB 1823419

References

1. White III RA, Rivas-Ubach A, Borkum MI, Köberl M, Bilbao A, Colby S. M., et al. The state of rhizospheric science in the era of multi-omics: a practical guide to omics technologies. 2017. Rhizosphere 3, 212-221.

2. Berendsen RL, Pieterse CM, Bakker PA. The rhizosphere microbiome and plant health. 2012.Trends Plant Sci., 17(8):478-486.

3. White III RA, Borkum MI, Rivas-Ubach A, Bilbao A, Wendler JP, Colby SM, Köberl M, Jansson C. From data to knowledge: the future of multi-omics data analysis for the rhizosphere. 2017. Rhizosphere 3:222-229.

4. Friesen ML. Widespread fitness alignment in the legume–rhizobium symbiosis. 2012. New Phytol., 194:1096-1111.

5. Martínez-Hidalgo P, Hirsch AM. 2017. The Nodule Microbiome: N_2-Fixing Rhizobia Do Not Live Alone. 1, 70-82.

6 Köberl M, White III RA, Erschen S, El-Arabi TF, Jansson JK, Berg G. Draft genome sequence of Paenibacillus polymyxa strain Mc5Re-14, an antagonistic Root endophyte of Matricaria chamomilla. 2015. Genome Announcements. 3(4):e00861-15

7. Ahkami AH, White III RA, Handakumbura PP, Jansson C. Rhizosphere engineering: Enhancing sustainable plant ecosystem productivity 2017. Rhizosphere 3, 233-243.

8. White III RA, Brown J, Colby S, Overall CC, Lee J, Zucker J, Glaesemann KR, Jansson C, Jansson JK. ATLAS (Automatic Tool for Local Assembly Structures) – a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. 2017. PeerJ Preprints 5: e2843v1.

9. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. 2017. PLoS Comput Biol., 13:e1005595.

10. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. 2015. Genome Res., 25(7):1043-55.

11. Seemann T. Prokka: rapid prokaryotic genome annotation. 2014. Bioinformatics. 30(14):2068-9.