Draft Genome Sequence of *Leifsonia poae* strain BS71, Isolated from a Drought Microcosm

Authors: Hanaa Ahmed[a], Kristen M. DeAngelis[b], Maureen A. Morrow[a#]

[a] Department of Biology, State University of New York at New Paltz, New Paltz, New York, USA

[b] Department of Microbiology, University of Massachusetts, Amherst, Massachusetts, USA

Draft Genome Sequence of *Leifsonia poae* strain BS71

#Address correspondence to Maureen A. Morrow, morrowm@newpaltz.edu.

We report the draft genome sequence of *Leifsonia poae* strain BS71. This bacterium was isolated from a low soil moisture content model soil microcosm inoculated with forest soil that had been subject to chronic warming.

Climate change is increasing temperature and drought conditions globally (2). To examine climate effects on soils, a field warming experiment was established in the Harvard Research Forest (HRF) in Petersham, MA (42.54°N, -72.18°W)(1). In October 2017, mineral horizon soil was collected from heated plots. A microcosm was generated by inoculating an 0.8 um filtered soil slurry to an artificial soil microcosm at 30% water content and incubated for 4 months at 15°C, fed weekly with cellobiose and NH_4NO_3, as described previously (3). These conditions were hypothesized to enrich for bacteria capable of growing in low soil moisture. BS71 was isolated from the microcosm on 1% glucose/ 0.4% potato infusion (Sigma-Aldrich, St. Louis, MO) agar plate, pH6, under aerobic conditions for 8 days at 25°C in the dark. BS71 was identified as *Leifsonia poae* by analyzing the 16S rRNA PCR product produced with the 27F/1492R primer pair(4) with IDTAXA(5).

BS71 DNA was prepared for sequencing by growing a single colony on 10% Tryptic Soy agar at 25°C for 7 days in the dark, scraping the biomass, and extracting the DNA using the Qiagen genomic DNA protocol (Qiagen, Valencia, CA). Whole genome sequencing was completed at the University of Massachusetts Medical School (UMMS) sequencing center. The DNA was sheared using a Biorupter (Diagenode, NJ) to a mean size of 20Kb. A PacBio SMRTbell library kit was used to construct the library that was sequenced on the PacBio RS II platform. The 150,292 raw reads, generated from a single cell, were filtered with the SMRT portal P-filter module (minimum subread length: 50 nucleotides, minimum polymerase quality: 75, and minimum polymerase read length: 50 nucleotides) and the resultant 69,059 filtered reads had a read N_{50} of 7302 bases.

The genome was assembled using sprai v0.9.9.23 (https://anaconda.org/bioconda/sprai) and Canu v1.5 (6). The final draft assembly contained 5 contigs (contig N_{50}, 3.98Mb) and is estimated as 98.99% complete and 0.063% contaminated using CheckM v1.0.18 (7) in KBase (8). Gene annotations were completed within JGI's Integrated Microbial Genomes (IMG) MGAP v4.16.5(9) with the gene calling program: Prodigal v2.6.3 (10, 11). Default parameters were used for all software except where noted. The genome is 4,144,138 bp (94.3x coverage), with a GC content of 67.84% and is predicted to encode for 3961 proteins, a single rRNA operon, and 45 tRNA genes.

A manually curated list of drought-associated genes was compared between BS71 and the 20 *Leifsonia* genomes with the greatest 16S rRNA gene homology within IMG's database. A greater number of beta-glucosidase (EC 3.2.1.21) annotated genes were present in BS71 (24 genes), as compared to 10 or fewer such annotated genes in the other *Leifsonia* (Table 1). Beta-glucosidase enzymes are diverse and play an important role in biomass conversion of recalcitrant carbon (12). The presumptive BS71 drought tolerance is also supported by the annotation of two distinct aquaporin genes, and genes for the production and transport of osmoprotectants. This genome supports the hypothesis that drought conditions characteristic of climate change may select for bacteria with drought-associated traits.

Data availability. The 16S PCR product sequence accession number is OL515151.The raw whole genome sequence reads are available in GenBank under the BioProject accession number

PRJNA745001. The Sequence Read Archive (SRA) accession number is SRR15142240 and the nucleotide sequence accession number is JAIHLP000000000. The annotation reported in this study is available at the Joint Genome Institute as Leifsonia poae BS71 first assembly.

Acknowledgements

This work was supported by the National Science Foundation CAREER award program and associated Research Opportunity Award (ROA) under contract DEB-1749206 (K.M.D. and M.A.M.), as well as the NSF Long-Term Ecological Research program (LTER) grant 1237491 to Harvard Forest. H.A. and M.A.M. would like to thank the SUNY New Paltz Research, Scholarship and Creative Activities Program for Summer Undergraduate Research Experience funding.

References

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.

2. Pokhrel Y, Felfelani F, Satoh Y, Boulange J, Burek P, Gädeke A, Gerten D, Gosling SN, Grillakis M, Gudmundsson L, Hanasaki N, Kim H, Koutroulis A, Liu J, Papadimitriou L, Schewe J, Müller Schmied H, Stacke T, Telteu C-E, Thiery W, Veldkamp T, Zhao F, Wada Y. 2021. Global terrestrial water storage and drought severity under climate change. Nature Climate Change 11:226–233.

3. Domeignoz-Horta LA, Pold G, Liu X-JA, Frey SD, Melillo JM, DeAngelis KM. 2020. Microbial diversity drives carbon use efficiency in a model soil. Nature Communications 11:3684.

4. Lane DJ. 1991. 16S/23S rRNA Sequencing, p. 115–175. *In* Nucleic Acid Techniques in Bacterial Systematic. John Wiley and Sons, Inc.

5. Murali A, Bhargava A, Wright ES. 2018. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome 6:140.

6. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 27:722–736.

7. 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, 2015/05/14 ed. 25:1043–1055.

8. 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. Nature Biotechnology 36:566–569.

9. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). Stand Genomic Sci 10:86–86.

10. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2018. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Research 47:D666–D677.

11. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

12. Ketudat Cairns JR, Esen A. 2010. β-Glucosidases. Cellular and Molecular Life Sciences 67:3389–3405.

Table 1 protein-protein BLAST results for beta-glucosidase (EC 3.2.1.21) annotated genes

Gene ID	Taxon identified by BLASTp highest score	percent identity	amino acids	IMG annotation description
2806535936	Humibacter sp. WJ7-1	71	846	glycoside hydrolase family 3 C-terminal domain-containing protein
2806536034	Streptacidiphilus fuscans	63	401	family 1 glycosylhydrolase
2806536039	Leifsonia shinshuensis	70	96	family 1 glycosylhydrolase
2806536324	Microbacterium azadirachtae	60	788	glycoside hydrolase family 3 C-terminal domain-containing protein
2806536495	Leifsonia shinshuensis	84	599	glycoside hydrolase family 3 protein
2806536706	Leifsonia sp. NCR5	81	831	glycoside hydrolase family 3 C-terminal domain-containing protein

2806536737	Frondihabitans sp. 762G35	73	776	glycoside hydrolase family 3 C-terminal domain-containing protein
2806536747	Leifsonia sp. Root227	94	391	family 1 glycosylhydrolase
2806536801	Leifsonia sp. NCR5	82	617	glycoside hydrolase family 3 C-terminal domain-containing protein
2806537199	Leifsonia sp. PS1209	79	501	beta-glucosidase
2806537254	Plantibacter flavus	65	592	glycoside hydrolase family 3 C-terminal domain-containing protein
2806537278	Microbacterium sp. Root61	91	761	glycoside hydrolase family 3 C-terminal domain-containing protein
2806538045	Rathayibacter sp. AY1A3	67	805	glycoside hydrolase family 3 C-terminal

				domain-containing protein
2806538050	Nonomuraea sp. 160415	61	403	family 1 glycosylhydrolase
2806538065	Thermocatellispora tengchongensis	70	746	glycoside hydrolase family 3 C-terminal domain-containing protein
2806538066	Microbacteriaceae bacterium	78	389	glycosyl hydrolase family protein
2806538078	Thermocatellispora tengchongensis	64	797	beta-glucosidase
2806538079	unclassified Leifsonia	77	609	MULTISPECIES: glycoside hydrolase family 3 C-terminal domain-containing protein
2806538223	Plantibacter sp. YR521	69	781	glycoside hydrolase family 3 C-terminal domain-containing protein

2806535250	Actinoplanes sp. OR16	68	578	glycoside hydrolase family 3 C-terminal domain-containing protein
2806535263	Streptomyces sp. yr375	83	409	family 1 glycosylhydrolase
2806535267	Leifsonia sp. Root227	88	751	ABC transporter substrate-binding protein
2806535268	Leifsonia sp. Root227	86	786	glycoside hydrolase family 3 C-terminal domain-containing protein
2806534677	Mycobacterium sp.	63	747	glycoside hydrolase family 3 C-terminal domain-containing protein