Genome Sequence of *Lysinibacillus sphaericus*, a Lignin-Degrading Bacterium Isolated from Municipal Solid Waste Soil

Gabriela F. Persinoti,a Douglas A. A. Paixão,a Timothy D. H. Bugg,b Fabio M. Squinac

aLaboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil
bDepartment of Chemistry, University of Warwick, Coventry, United Kingdom
cPrograma de Processo Tecnológicos e Ambientais da Universidade de Sorocaba (UNISO), Sorocaba, Brazil

ABSTRACT We report here the draft genome sequence of *Lysinibacillus sphaericus* strain A1, a potential lignin-degrading bacterium isolated from municipal solid waste (MSW) soil and capable of enhancing gas release from lignocellulose-containing soil.

*L*ysinibacillus sphaericus strain A1 is a Gram-positive bacterium from the *Firmicutes* phylum that was isolated from municipal solid waste (MSW) soil in the United Kingdom (1). This lignin-degrading bacterium strain is particularly interesting since, under microscale anaerobic conditions, it presents enhanced methane release from lignocellulose-containing soil, suggesting the potential for *in situ* treatment and enhancement of landfill soil gas production (1).

We report here the genome sequence of *L. sphaericus* strain A1. We sequenced the genome of this strain on an Illumina HiSeq 2500 system at the CTBE next-generation sequencing (NGS) facility, generating 5,262,224 paired-end reads (insert size, 300 bp) and 4,454,143 mate pair reads (3 libraries with an insert size between 3 and 4 kb, 5 and 7 kb, and 8 and 11 kb). Paired-end reads were preprocessed with Trimmomatic (2) to remove the adapter and low-quality sequences, and mate pair reads were processed using NextClip (3), resulting in 5,178,788 and 1,227,816 high-quality reads, respectively.

The genome size of *L. sphaericus* was estimated to be 4.54 Mb based on k-mer count statistics accessed with KmerGenie (4), with an estimated coverage of approximately 200×. Genome assembly was carried out with SPAdes v3.6.2 (5) using an assembly of k-mer values (21, 33, 47, 55, and 77) and SSPACE (6). The presence of typical bacterial marker genes in the assembled genome was assessed using CheckM (7), which estimated the completeness of the genome to be 99.91%.

The resulting assembly for *L. sphaericus* has nine scaffolds with a total length of 4,517,188 bp and an N_{50} value of 3,305,581 bp. The average GC content of the genome is 37%.

Gene prediction and annotation were carried out using the Prokka prokaryotic genome annotation pipeline (8). A total of 4,371 genes were identified in the *L. sphaericus* genome; of these, there are 4,305 protein-encoding genes, 9 rRNAs, 56 tRNAs, and 1 transfer-messenger RNA (tmRNA). Regarding 16S rRNA, *L. sphaericus* strain A1 presents 99.92% sequence similarity with *L. sphaericus* KCTC 3346T, 99.76% with *L. fusiformis* NBRC 15717T, and 99.60% with *L. mangiferihumi* M-GX18T, as analyzed using the EzBioCloud Web server identity tool (9).

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number PGLV00000000. The version described in this paper is version PGLV01000000.

Address correspondence to Fabio M. Squina, fabio.squina@prof.uniso.br.

REFERENCES

1. Rashid GMM, Durán-Peña MJ, Rahmanpour R, Sapsford D, Bugg TDH. 2017. Delignification and enhanced gas release from soil containing lignocellulose by treatment with bacterial lignin degraders. J Appl Microbiol 123:159–171. https://doi.org/10.1111/jam.13470.
2. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
3. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics 30:566–568. https://doi.org/10.1093/bioinformatics/btt702.
4. Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30:31. https://doi.org/10.1093/bioinformatics/btt310.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
6. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579. https://doi.org/10.1093/bioinformatics/btq683.
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 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
9. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.