Complete Genome Sequence of *Bacillus pumilus* PDSLzg-1, a Hydrocarbon-Degrading Bacterium Isolated from Oil-Contaminated Soil in China

Kun Hao,[b] Hongna Li,[a] Feng Li,[a] Ping Guo[a]

Agricultural Clean Watershed Research Group, Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences, Beijing, People's Republic of China[a]; State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, People's Republic of China[b]

Bacillus pumilus strain PDSLzg-1, an efficient hydrocarbon-degrading bacterium, was isolated from oil-contaminated soil. Here, we present the complete sequence of its circular chromosome and circular plasmid. The genomic information is essential for the study of degradation of oil by *B. pumilus* PDSLzg-1.

Address correspondence to Ping Guo, guoping@caas.cn.

Bacillus pumilus strain PDSLzg-1 is a novel bacterium isolated from oil-contaminated soil of the Shengli oil field (Shandong, People's Republic of China) in 2015. The bacterium is characterized as Gram-positive, aerobic, motile, rod-shaped, and forming endospores. According to 16s rRNA sequencing analysis, the bacterium is identified as a strain of *Bacillus pumilus*. This strain, designated PDSLzg-1, has the ability to effectively degrade petroleum hydrocarbons, especially C8 to C18 alkanes, and was shown to produce at least one bio-surfactant and may be of interest for an industrial application of oil pollution degradation.

B. pumilus PDSLzg-1 genomic DNA was isolated from overnight culture using a Wizard Genomic DNA purification kit (Promega). Total DNA obtained was subjected to quality control by agarose gel electrophoresis and quantified by Qubit. The genome of *B. pumilus* PDSLzg-1 was sequenced by single molecule real-time (SMRT) technology. Sequencing was performed at the Beijing Novogene Bioinformatics Technology Co., Ltd., producing 95,240 reads with 358× coverage. SMRT Analysis 2.3.0 was used to filter low-quality reads and the filtered reads were assembled to generate one contig without gaps. The genome of *B. pumilus* PDSLzg-1 had a 3,698,973-bp circular chromosome and an 11,801-bp circular plasmid with G+C contents of 41.96% and 39.36%, respectively. Gene prediction was performed on the *B. pumilus* PDSLzg-1 genome by GeneMarkS, resulting in the identification of 3,895 genes with an average length of 848 bp (1). 81 tRNA genes were predicted with tRNAscan-SE, 24 rRNA genes were predicted with rRNAmmer, and seven sRNAs were predicted by a BLAST search against the Rfam database (2–4). A whole-genome BLAST search (*E* value less than 1e−5, minimal alignment length percentage larger than 40%) was performed against KEGG, COG, nr, Swiss-Prot, GO, and TrEMBL (5–9). 3,689 (94.7%) genes were annotated to at least one database.

There were a group of genes in genome of *B. pumilus* strain PDSLzg-1 encoding alcohol dehydrogenase, aldehyde dehydrogenase, and monooxygenase, which are probably related to oil degradation (10, 11). This study presents the first fully sequenced and annotated hydrocarbon-degrading strain of *B. pumilus*, which is valuable for studying its evolution, hydrocarbon-degrading mechanism, and industrial applications.

Accession number(s). The complete annotated genome and plasmid sequences of *B. pumilus* strain PDSLzg-1 were deposited in GenBank under the accession numbers CP016784 and CP016785, respectively.

ACKNOWLEDGMENTS

REFERENCES

1. **Besemer J, Lomsadze A, Borodovsky M**. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res **29**: 2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.
2. **Lowe TM, Eddy SR**. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25**: 955–964. http://dx.doi.org/10.1093/nar/25.5.0955.
3. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW**. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35**:3100−3108. http://dx.doi.org/10.1093/nar/gkm160.
4. **Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A**. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res **37**: D136−D140. http://dx.doi.org/10.1093/nar/gkn766.
5. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. 1990. Basic local alignment search tool. J Mol Biol **215**:403−410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
6. **Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M**. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res **32**: D277−D280. http://dx.doi.org/10.1093/nar/gkh063.
7. **Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS,**

Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**:41. http://dx.doi.org/10.1186/1471-2105-4-41.

8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. Nat Genet **25**:25–29. http://dx.doi.org/10.1038/75556.

9. Magrane M, UniProt Consortium. 2011. UniProt Knowledgebase: a hub of integrated protein data. Database **2011**:bar009. http://dx.doi.org/10.1093/database/bar009.

10. Kotani T, Yamamoto T, Yurimoto H, Sakai Y, Kato N. 2003. Propane monooxygenase and NAD$^+$-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. J Bacteriol **185**: 7120–7128. http://dx.doi.org/10.1128/JB.185.24.7120-7128.2003.

11. Harayama S, Kishira H, Kasai Y, Shutsubo K. 1999. Petroleum biodegradation in marine environments. J Mol Microbiol Biotechnol **1**:63–70.