Draft Genome Sequence of a Fermenting Bacterium, *Soehngenia* sp. Strain 1933P, Isolated from a Petroleum Reservoir in Azerbaijan

Denis S. Grouzdev,[a] Salimat K. Bidzhieva,[b] Diyana S. Sokolova,[b] Tatiyana P. Tourova,[b] Andrey B. Poltaraus,[c] Tamara N. Nazina[b]

[a]Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russian Federation
[b]Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russian Federation
[c]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russian Federation

ABSTRACT The draft genome sequence of a mesophilic fermenting bacterium, *Soehngenia* sp. strain 1933P, isolated from production water of the Binagady petroleum reservoir (Republic of Azerbaijan), is presented. The genome is annotated for elucidation of the metabolic potential and taxonomic position of strain 1933P.

The mesophilic fermenting bacterium *Soehngenia* sp. strain 1933P (VKM B-3382) was isolated from a methanogenic enrichment culture obtained from the Binagady petroleum reservoir (Republic of Azerbaijan) (1). The methanogenic enrichment was cultivated in the mineral medium with methanol (5% [vol/vol]) and 500 mg $Na_2S \cdot 9H_2O$ (2) and contained fermentative bacteria and methanogens of the genus *Methanosarcina* (3). Several flasks with the methanogenic enrichment were stored at room temperature (18 to 24°C) for 33 years without transfers to fresh medium for studying the survival of oilfield communities. Then, the enrichment was transferred to fresh medium with methanol, and a pure culture of strain 1933P was isolated by sequential transfers from the highest dilutions on the mineral medium (2) with peptone (2 g liter^{-1}), yeast extract (0.2 g liter^{-1}), 1.5% NaCl, and $Na_2S \cdot 9H_2O$ (0.2 g $\cdot$ liter^{-1}) at 30°C. To ascertain the culture purity, material from the highest growth-positive dilution (10^{-7}) was used to inoculate the fresh medium with peptone. After 5 to 7 days of incubation at 30°C, the 16S rRNA gene was amplified with the 27F and 1492R primers (4), and purified PCR products were sequenced with an ABI Prism 3730 DNA analyzer (Applied Biosystems, USA). The 16S rRNA sequence analysis using a BLASTn (5) search against the NCBI database revealed that strain 1933P was phylogenetically closely related (98.5% similarity) to *Soehngenia saccharolytica* DSM 12858^T (6), a strain from the only species of the genus *Soehngenia*. The aim of the present study was to sequence the genome of the strain 1933P in order to elucidate its metabolic potential and taxonomic position.

Strain 1933P was grown anaerobically at 30°C in the mineral medium (2) used for the isolation of pure culture. Cells were harvested from 2 liters of culture medium by centrifugation after 7 days of incubation. DNA was purified from the cell biomass using the cetyltrimethylammonium bromide (CTAB) method (7). The libraries were constructed with the NEBNext DNA library prep reagent set for Illumina, according to the protocol for the kit. Sequencing of genomic DNA was carried out using the HiSeq 2500 platform (Illumina, Inc., USA), with 150-bp paired-end reads. A total of 4,678,392 reads were obtained from strain 1933P. Raw sequence reads were quality checked with FastQC version 11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and low-quality reads were trimmed using Trimmomatic version 0.36 (8), with the default settings for paired-end reads. Subsequently, the quality-filtered reads were

de novo assembled with SPAdes version 3.13.0 using the default settings (9). The final assembled 1,917,091-bp-long genome comprised 33 scaffolds, with an N_{50} value of 132,646 bp, G+C content of 31.9%, and coverage of 630×. The average nucleotide identity (ANI) (10) and digital DNA-DNA hybridization (dDDH) (http://ggdc.dsmz.de/ggdc.php) (11) values of 83.5% and 27.0%, respectively, to the genome of the most closely related species, *S. saccharolytica* DSM 12858^T, were below the species cutoffs (95 to 96% for ANI and 70% for dDDH) (12), which indicates that strain 1933P belongs to a new *Soehngenia* species.

Identification of protein-coding sequences and primary annotation were performed using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (13). The draft genome sequence of *Soehngenia* sp. 1933P contained 1,853 genes, of which 1,789 were protein-coding sequences, 23 were pseudogenes, and 41 coded for RNAs. Functional annotation of the genome performed with the Rapid Annotations using Subsystems Technology (RAST) server (14, 15), via the RAST*tk* pipeline with the default settings (16), revealed that 150 of the genes were associated with protein metabolism, 106 genes were associated with the metabolism of amino acids and derivatives, 101 genes were associated with carbohydrate metabolism, and 40 genes were associated with the metabolism of cofactors, vitamins, prosthetic groups, and pigments. The genome sequence of *Soehngenia* sp. 1933P provided here will broaden the knowledge of the genus *Soehngenia* and the putative importance of its members in subsurface communities.

Data availability. The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of *Soehngenia* sp. strain 1933P is MK712484. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number SRIB00000000. The version described in this paper is version SRIB01000000. The associated BioProject, BioSample, and SRA accession numbers are PRJNA529817, SAMN11286712, and SRR8846858, respectively. RAST annotation of *Soehngenia* sp. 1933P is publicly available using the guest account for the RAST online database (job ID 707512).

REFERENCES

1. Nazina TN, Rozanova EP, Kuznetsov SI. 1985. Microbial oil transformation processes accompanied by methane and hydrogen-sulfide formation. Geomicrobiol J 4:103–130. https://doi.org/10.1080/01490458509385927.
2. Zeikus JG, Weimer PJ, Nelson DR, Daniels L. 1975. Bacterial methanogenesis: acetate as a methane precursor in pure culture. Arch Microbiol 104:129–134. https://doi.org/10.1007/BF00447312.
3. Nazina TN. 1984. Communities of methane-producing bacteria from Apsheron oil formations. Microbiology (Moscow) 53:122–127.
4. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A 82:6955–6959. https://doi.org/10.1073/pnas.82.20.6955.
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
6. Parshina SN, Kleerebezem R, Sans JL, Lettinga G, Nozhevnikova AN, Kostrikina NA, Lysenko AM, Stams A. 2003. *Soehngenia saccharolytica* gen. nov., sp. nov., and *Clostridium amygdalinum* sp. nov., two novel, anaerobic, benzaldehyde-converting bacteria. Int J Syst Evol Microbiol 53:1791–1799. https://doi.org/10.1099/ijs.0.02668-0.
7. Wilson K. 2001. Preparation of genomic DNA from bacteria. Curr Protoc Mol Biol Chapter 2:Unit 2.4. https://doi.org/10.1002/0471142727.mb0204s56.
8. 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.
9. 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.
10. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.
11. Auch AF, von Jan M, Klenk HP, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2:117–134. https://doi.org/10.4056/sigs.531120.
12. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes.

Int J Syst Evol Microbiol 68:461–466. https://doi.org/10.1099/ijsem.0 .002516.

13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W. 2013. Prokaryotic Genome Annotation Pipeline. The NCBI handbook, 2nd ed. National Center for Biotechnology Information, Bethesda, MD.

14. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

15. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

16. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.