Originally published as:

Genome Sequence of *Methanosarcina soligelidi* SMA-21, Isolated from Siberian Permafrost-Affected Soil

Mashal Alawi,[a] Nicole Shapiro,[b] Tanja Woyke,[b] Fabian Horn,[a] Corien Bakermans,[c] Dirk Wagner[a]

GFZ German Research Centre for Geosciences, Section 4.5 Geomicrobiology, Potsdam, Germany[a]; DOE Joint Genome Institute, Walnut Creek, California, USA[b]; Altoona College, The Pennsylvania State University, Altoona, Pennsylvania, USA[c]

Here, we announce the genome sequence of *Methanosarcina soligelidi* SMA-21, an anaerobic methanogenic archaeon that was previously isolated from Siberian permafrost-affected soil. The sequencing of strain SMA-21 yielded a 4.06-Mb genome with 41.5% G+C content, containing a total of 2,647 open reading frames.

Address correspondence to Mashal Alawi, mashal.alawi@gfz-potsdam.de.

*M*ethanosarcina soligelidi (type strain SMA-21 = DSM 26065 = JCM 18468) was initially isolated from Siberian permafrost-affected soil and described as a novel species within the order *Methanosarcinales* ([1]). The 16S rRNA gene sequence (NCBI AB973359) was close to that of *Methanosarcina mazei* (99.9%). The strain grows on H_2/CO_2, methanol, and acetate and has a high survival potential against air exposure, desiccation, freeze-thaw cycles, and long-term freezing.

Here, we report the complete genome sequence of *M. soligelidi* SMA-21. The genome was generated at the Department of Energy (DOE) Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology ([2]). A PacBio SMRTbell library was constructed and sequenced on the PacBio RS platform, which generated 164,129 filtered subreads totaling 551.7 Mbp. The raw reads were assembled using HGAP (version 2.1.1) ([3]). The final draft assembly contained 1 contig in 1 scaffold, totaling 4.1 Mbp in size (G+C content, 41.50%). Genes were identified using Prodigal ([4]), followed by a round of manual curation using GenePRIMP ([5]). The tRNAscan-SE tool ([6]) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA ([7]). Other noncoding RNAs, such as the RNA components of the protein secretion complex and RNase P, were identified by searching the genome for the corresponding Rfam profiles using Infernal (http://infernal.janelia.org). Additional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes (IMG) platform ([8]) developed by JGI, Walnut Creek, CA, USA (http://img.jgi.doe.gov). The genome contains 3,440 protein-coding sequences, of which 75.4% (2,647) had a predicted function. Likewise, the genome contains five clustered regularly interspaced short palindrome repeat (CRISPR) loci and CRISPR-associated proteins (Cas). The genetic, metabolic, and physiological features of the species will be unveiled by future comparative genomic analyses.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank un-

der the accession no. JQLR00000000. The version described in this paper is the first version.

REFERENCES

1. Wagner D, Schirmack J, Ganzert L, Morozova D, Mangelsdorf K. 2013. *Methanosarcina soligelidi* sp. nov., a desiccation- and freeze-thaw-resistant methanogenic archaeon from a Siberian permafrost-affected soil. Int J Syst Evol Microbiol **63**:2986–2991. http://dx.doi.org/10.1099/ijs.0.046565-0.
2. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D. 2009. Real-time DNA sequencing from single polymerase molecules. Science **323**:133–138. http://dx.doi.org/10.1126/science.1162986.
3. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10**:563–569. http://dx.doi.org/10.1038/nmeth.2474.
4. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. http://dx.doi.org/10.1186/1471-2105-11-119.
5. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lyki-

dis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods 7:455–457. http://dx.doi.org/10.1038/nmeth.1457.

6. 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.

7. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. Silva: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196. http://dx.doi.org/10.1093/nar/gkm864.

8. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics 25:2271–2278. http://dx.doi.org/10.1093/bioinformatics/btp393.