Draft Genome Sequence of *Methanobacterium paludis* IBT-C12, Recovered from Sediments of the Apatlaco River, Mexico

Erick Alejandro Sánchez-Salazar,ᵃ Lizbeth Hernández-Jaimes,ᵇ Luz Breton-Deval,ᶜ ◉ Ayixon Sánchez-Reyesᶜ

ᵃUniversidad Autónoma Metropolitana, Unidad Iztapalapa, Mexico City, Mexico
ᵇCentro de Investigación en Dinámica Celular, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico
ᶜCátedras Conacyt-Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

ABSTRACT *Methanobacterium paludis* is a hydrogenotrophic archaea first described in 2014 and isolated from a peatland area. So far, there is only one sequenced genome of this taxon. Here, we report the draft genome sequence of *M. paludis* IBT-C12, a metagenome-assembled genome (MAG) from sediments in the Apatlaco River, Mexico.

M *ethanobacterium paludis* is an anaerobic archaeon initially described in 2014 and isolated from a peatland area. It uses a hydrogenotrophic pathway to reduce CO_2 to methane on methanogenesis (1). There is little information on the genomic complement of this methanogenic species, with only one available assembly for the type strain, SWAN1, on the National Center for Biotechnology Information (NCBI) (accession number GCA_000214725.1) (2). Other genomic composites would be crucial to gain information on the metabolism, ecology, taxonomy, and other traits of environmental relevance. Here, we report the draft genome sequence of *M. paludis* IBT-C12, a metagenome-assembled genome (MAG) recovered from freshwater sediments in a highly polluted river in Mexico (3, 4). The MAG was isolated from an environmental sample. Full information about the sample nature, collection site, deconvolution processes, and data were reported before (3). Briefly, sediments were collected at 10-cm depth at a rate of 0.5 kg per site and were stored in coolers at 4°C until further processing. Metagenomic DNA was extracted using the DNeasy PowerWater kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. The *M. paludis* IBT-C12 genome was deconvolved from two sets of short reads obtained from freshwater and sediment in the Apatlaco River, Mexico. One read set was acquired from freshwater on a shotgun NextSeq 500 sequencing platform (Illumina, Inc., San Diego, CA) (108,785,988 paired-end reads, 75 bp) and a set of Hi-C reads from sediments (72,007,031 paired-end reads, 150 bp) from a HiSeq 4000 Illumina-compatible sequencing library. The shotgun library was prepared using the TruSeq kit version 2 (Illumina, Inc.) according the manufacturer's protocol. The ProxiMeta microbiome kit (Phase Genomics, Seattle, WA) was used to prepare the Hi-C library, and the DNA was digested with Sau3AI and MluCI enzymes before proximity ligation. Read quality was assessed and filtered under the Phase Genomics cloud-based bioinformatics portal by using the Fastp tool version 0.23.1 (5, 6). A metagenome assembly was constructed with the shotgun reads (only the NextSeq data were used) using MEGAHIT software version 1.2.9 (7) (size 550,428,717 bp and 746,031 contigs). The Hi-C reads were subsequently mapped with the Burrows-Wheeler Aligner MEM algorithm (BWA-MEM) version 0.7.17 (-5SP and -t 8 options) to create Hi-C-based contact probability maps (8). The binning was performed with ProxiMeta software (5), resulting in 97 MAGs (3). The MAG corresponding to IBT-C12 was identified by comparing genome relatedness indexes (mash distance [9] and average nucleotide identity [ANI] [10, 11]) against a custom database (12). Default parameters were used for all software unless otherwise noted. The total

TABLE 1 Average nucleotide identity and Mash mutational distance (D) of *Methanobacterium paludis* IBT-C12 compared with the closest NCBI type strains of the genus *Methanobacterium*

GenBank assembly accession no.	Reference organism	OrthoANI (%)	Mash D
GCA_000214725.1	*Methanobacterium paludis* SWAN1	95.20	0.05
GCA_017874455.1	*Methanobacterium aggregans* DSM 29428	78.02	0.22
GCA_002287175.1	*Methanobacterium bryantii* M.o.H	71.07	0.26
GCA_000745485.1	*Methanobacterium veterum* MK4	70.64	0.26
GCA_000746075.1	*Methanobacterium arcticum* M2	70.60	0.26
GCA_001316325.1	*Methanobacterium formicicum* JCM 10132	69.41	0.30
GCA_017873625.1	*Methanobacterium petrolearium* DSM 22353	74.38	0.30
GCA_002813695.1	*Methanobacterium subterraneum* A8p	69.33	1.00
GCA_000016525.1	*Methanobrevibacter smithii* DSMZ 861	65.91	1.00

length of the *M. paludis* IBT-C12 genome was 2,696,644 bp, with a G+C content of 35.95%. The number of contigs was 374, with an average shotgun sequencing coverage depth of 16× and an N_{50} value of 1,849,042 bp. The completeness and contamination were estimated with miComplete version 1.1.1 (13) at 61.83% and 1.13%, respectively. The genome size was close to that of the reference strain, SWAN1 (2,546,541 bp), so IBT-C12 MAG must be moderately complete.

The IBT-C12 strain is closely related to *M. paludis* SWAN1, as shown by a FastANI result of 94.34% (95.2% with OrthoANI version 0.5.0) and the phylogenomic analysis (Table 1; Fig. 1) (10, 14). The annotation of the MAG sequence was performed with the DFAST service version web with default parameters (15). A total of 2,550 genes were identified, with 2,514 protein-coding sequences and 36 RNA genes. The RNA genes comprised 1 partial 16S rRNA, 2 partial 23S rRNAs, and 33 tRNAs; 2 CRISPR arrays were also identified.

The MAG described in this work could provide valuable information regarding the ecology, metabolism, phylogeny, and evolution of the *M. paludis* clade.

FIG 1 FastTree phylogenomic tree using 25 core genes inferred with the UBCG pipeline (14, 16). The 25 core genes were automatically aligned with MAFFT and concatenated (17). Poorly conserved regions were curated with Gblocks (18). Gene support indices (GSIs) as percentages are shown at branching points. Bars indicate substitution per site.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the version number JAIPRP000000000.1. The version described in this paper is version JAIPRP000000000.2. The BioProject accession number is PRJNA759916. The BioSample accession number is SAMN21209357. The draft shotgun metagenome assembly is available on https://figshare.com/ndownloader/files/28075893. The Hi-C reads and NextSeq data are available under accession numbers SRR11481801 and SRX6045636 to SRX6045639, respectively, in the SRA.

REFERENCES

1. Cadillo-Quiroz H, Brauer SL, Goodson N, Yavitt JB, Zinder SH. 2014. Methanobacterium paludis sp. nov. and a novel strain of Methanobacterium lacus isolated from northern peatlands. Int J Syst Evol Microbiol 64:1473–1480. https://doi.org/10.1099/ijs.0.059964-0.

2. Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, Smith RG, Tatusova T, Xiang C, Zherikov A, DiCuccio M, Murphy TD, Pruitt KD, Kimchi A. 2016. Assembly: a resource for assembled genomes at NCBI. Nucleic Acids Res 44:D73–D80. https://doi.org/10.1093/nar/gkv1226.

3. Sánchez-Reyes A, Bretón-Deval L, Mangelson H, Salinas-Peralta I, Sanchez-Flores A. 2021. Hi-C deconvolution of a textile dye–related microbiome reveals novel taxonomic landscapes and links phenotypic potential to individual genomes. Int Microbiol 25:99–110. https://doi.org/10.1007/s10123-021-00189-7.

4. Breton-Deval L, Sanchez-Reyes A, Sanchez-Flores A, Juárez K, Salinas-Peralta I, Mussali-Galante P. 2020. Functional analysis of a polluted river microbiome reveals a metabolic potential for bioremediation. Microorganisms 8:554. https://doi.org/10.3390/microorganisms8040554.

5. Press MO, Wiser AH, Kronenberg ZN, Langford KW, Shakya M, Lo C, Mueller KA, Sullivan ST, Chain PSG, Liachko I. 2017. Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. bioRxiv https://doi.org/10.1101/198713.

6. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

7. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

8. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

9. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:1–14. https://doi.org/10.1186/s13059-016-0997-x.

10. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.

11. Lee I, Kim YO, Park SC, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760.

12. Sánchez-Reyes A, Fernández-López MG. 2021. Mash sketched reference dataset for genome-based taxonomy and comparative genomics. Preprints https://doi.org/10.20944/PREPRINTS202106.0368.V1.

13. Hugoson E, Lam WT, Guy L. 2019. miComplete: weighted quality evaluation of assembled microbial genomes. Bioinformatics 36:936–937. https://doi.org/10.1093/bioinformatics/btz664.

14. Na SI, Kim YO, Yoon SH, Ha S-M, Baek I, Chun J. 2018. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. J Microbiol 56:281–285. https://doi.org/10.1007/s12275-018-8014-6.

15. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039. https://doi.org/10.1093/bioinformatics/btx713.

16. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

17. Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286–298. https://doi.org/10.1093/bib/bbn013.

18. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564–577. https://doi.org/10.1080/10635150701472164.