

Complete Genome Sequence of a Nonculturable *Methanococcus maripaludis* Strain Extracted in a Metagenomic Survey of Petroleum Reservoir Fluids

Xiaoyi Wang,¹ Paul Greenfield,² Dongmei Li,³ Philip Hendry,^{3*}
Herbert Volk,⁴ and Tara D. Sutherland¹

CSIRO Ecosystem Sciences, P.O. Box 1700, Acton, ACT 2601, Australia¹; CSIRO Mathematics Informatics and Statistics, Locked Bag 17, North Ryde, NSW 1670, Australia²; CSIRO Food and Nutritional Sciences, P.O. Box 52, North Ryde, NSW 1670, Australia³; and CSIRO Earth Science and Resource Engineering, P.O. Box 132, North Ryde, NSW 1670, Australia⁴

Received 17 July 2011/Accepted 25 July 2011

Extraction of genome sequences from metagenomic data is crucial for reconstructing the metabolism of microbial communities that cannot be mimicked in the laboratory. A complete *Methanococcus maripaludis* genome was generated from metagenomic data derived from a thermophilic subsurface oil reservoir. *M. maripaludis* is a hydrogenotrophic methanogenic species that is common in mesophilic saline environments. Comparison of the genome from the thermophilic, subsurface environment with the genome of the type species will provide insight into the adaptation of a methanogenic genome to an oil reservoir environment.

A complete genome of a new *Methanococcus maripaludis* strain (X1) was isolated from metagenomic data generated from a subsurface thermophilic saline oil reservoir. *M. maripaludis* is a hydrogenotrophic methanogen widely found in mesophilic anaerobic saline surface environments (1–3) and recently noticed by petroleum microbiologists because of its ability to corrode petroleum reservoir surface infrastructure (3, 4). Despite the abundance of *M. maripaludis* X1 genomic DNA in our metagenomic data pool, we could not detect *M. maripaludis* X1 16S rRNA genes in laboratory cultures incubated under reservoir conditions after the addition of a variety of nutrients, consistent with the recent finding that environmental organisms that are abundant and cosmopolitan are less likely to grow rapidly under rich nutrient conditions (5). To our knowledge, the *M. maripaludis* X1 genome is the first genome from a noncultured microorganism reconstructed directly from *de novo* sequencing of a metagenomic data pool.

Formation water was collected from an offshore oil field (approximately 800-m depth, 50°C, with ~2.3% salinity) near Malaysia that had not been subject to water flooding or other treatments. Total DNA was extracted from 6 liters of fluid using the WaterMaster DNA purification kit (Epicentre, Madison, WI), and sequencing was performed using a combination of GS-FLX and Illumina (Ramaciotti Centre, Sydney, Australia) sequencing technologies.

The GS-FLX reads were initially assembled using Newbler (454 Life Sciences), and then both the GS-FLX and the paired-end Illumina data were assembled using Velvet (6) to produce 7,719 contigs from the complete set of metagenomic reads. Alignment of all contigs to the *M. maripaludis* S2 reference genome produced a subset of 42 that covered most of the S2 reference genome. Five of these contigs were >100 kbp in length, and 11 were over 50 kbp, with the longest being 153,187

bp. These contigs (with an average 6-fold coverage from GS-FLX reads and 101-fold coverage from Illumina reads) formed the skeleton (total of 1,611,868 bp) used to guide the assembly of the complete *M. maripaludis* X1 genome. The lengths of the contigs indicated that each was derived from a single strain, as the presence of DNA from multiple strains would have caused Velvet to break them into small pieces at regions of significant difference. In fact, point differences observed at the ends of some contigs, possibly due to slight variations in the genome of the X1 organism or sequencing errors, caused Velvet to create multiple contigs. The gaps between contigs and ambiguous patches within contigs (i.e., NNNNNNNNNN fragments; about 7.7% of the total genome) were filled by manually aligning individual reads to the contig ends using in-house programs. The complete genome was annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline. It contained 1,746,697 bp with 32% G+C content, 1,892 predicted genes, and 2 rRNA operons.

Nucleotide sequence accession number. The complete genome sequence for the *M. maripaludis* X1 genome was deposited under GenBank accession no. CP002913.

This study was supported by CSIRO and Petronas.

REFERENCES

1. Hendrickson, E. L., et al. 2004. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J. Bacteriol.* **186**:6956–6969.
2. Jones, W. J., M. J. B. Paynter, and R. Gupta. 1983. Characterization of *Methanococcus maripaludis* sp. nov., a new methanogen isolated from salt marsh sediment. *Arch. Microbiol.* **135**:91–97.
3. Mori, K., H. Tsurumaru, and S. Harayama. 2010. Iron corrosion activity of anaerobic hydrogen-consuming microorganisms isolated from oil facilities. *J. Biosci. Bioeng.* **110**:426–430.
4. Uchiyama, T., K. Ito, K. Mori, H. Tsurumaru, and S. Harayama. 2010. Iron-corroding methanogen isolated from a crude-oil storage tank. *Appl. Environ. Microbiol.* **76**:1783–1788.
5. Yooseph, S., et al. 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**:60–67.
6. Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.

* Corresponding author. Mailing address: CSIRO Food and Nutritional Sciences, P.O. Box 52, North Ryde, NSW 1670, Australia. Phone: 61-2-94905099. Fax: 61-2-94905010. E-mail: Phil.Hendry@csiro.au.