genomeAnnouncements™

The Completed PacBio Single-Molecule Real-Time Sequence of *Methylosinus trichosporium* Strain OB3b Reveals the Presence of a Third Large Plasmid

John R. Heil,[a,b] Michael D. J. Lynch,[a,b] Jiujun Cheng,[a,b] Ola Matysiakiewicz,[a] Maya D'Alessio,[a] Trevor C. Charles[a,b]

Department of Biology, University of Waterloo, Waterloo, Ontario, Canada[a]; Metagenom Bio Inc., Toronto, Ontario, Canada[b]

ABSTRACT   Presented here is the complete genome sequence of the well-studied *Rhizobiales* methanotroph *Methylosinus trichosporium* strain OB3b. The assembly contains 5,183,433 bp, corresponding to a chromosome of 4,508,832 bp and three circular plasmids of 285,280 bp, 209,102 bp, and 180,219 bp.

Methylosinus trichosporium strain OB3b (1), a member of the *Methylocystaceae* family within the *Rhizobiales* order of *Alphaproteobacteria*, is an obligately aerobic methanotroph that has served as a model for the study of methanotrophy and the related ability to degrade hydrocarbons (2). A draft genome sequence, reported in 2010 (3), was the first one of a member of the *Methylocystaceae* family. This sequence included a single circular chromosome and two large plasmids. Transcriptomic and metabolomic studies have been facilitated by this draft sequence (4, 5).

In our quest to develop genome engineering methods for strain OB3b, we recognized the limitations brought on by lack of a completed genome sequence. We therefore decided to have the genome completed using PacBio single-molecule real-time (SMRT) long-read technology. A culture of strain OB3b was a gift of Jeremy Semrau. Colonies were grown on nitrate mineral salts (NMS) agar incubated at 30°C in a methane-air atmosphere. DNA was isolated from pooled colonies by using a cetyltrimethylammonium bromide (CTAB)-phenol-chloroform extraction method (6). The method was modified slightly; the SDS concentration during lysis was increased to 1%, and a chloroform extraction step was added after the phenol-chloroform extraction. The DNA was provided to McGill University and the Génome Québec Innovation Centre, where a sheared large-insert library was prepared and sequenced on the Pacific Biosciences RS-II SMRT instrument. The subread count was 132,509, with a mean length of 7,199, for a total of 953,876,416 bases, giving approximately 180-fold coverage. The raw-read quality score was 85.

Contigs were assembled by the sequencing facility using the Hierarchical Genome Assembly Process (HGAP) workflow. Briefly, raw subreads were extracted from the bax.h5 data files, and a preassembly was generated using BLASR (7). This preassembly then seeded the Celera assembler (8). Raw reads were aligned to the resulting contigs using BLASR, and high-quality consensus sequences were generated through Quiver variant calling. The assembly was circularized by the Circlator program (9) using the raw reads corrected by Canu v.1.6 (10). One unitig (corresponding to GenBank accession number CP023739) was not initially circularized despite an apparent 15-kb repeat at each end. A manual "break" was introduced roughly at the midpoint, and the sequences were reassembled using the toAmos and minimus2 (11) programs. The resulting unitig was returned to the original assembly, and Circlator was rerun, resulting

in four circular contigs, with a complete genome size of 5,183,433 bp. One contig of 4,508,832 bp corresponded to the chromosomal sequence in the previously determined draft genome but comprises four segments that are joined in an order and orientation that are different than those of the draft sequence. Two of the smaller contigs, 285,280 bp and 180,219 bp, corresponded to the two previously reported plasmids. The remaining contig appears to be a third large plasmid, of 209,102 bp, that was not reported in the draft sequence. Interestingly, sequence reads mapping to this plasmid were not found in the draft sequence or the reads contributing to the draft sequence. The culture whose DNA contributed to the draft sequence was likely lacking this plasmid. Each of the three plasmids contains an iconic *repABC* operon (12).

Accession number(s). The genome sequence reported here has been deposited in GenBank under the accession numbers CP023737, CP023738, CP023739, and CP023740 and is a component of BioProject number PRJNA413061.

REFERENCES

1. Whittenbury R, Phillips KC, Wilkinson JF. 1970. Enrichment, isolation and some properties of methane-utilizing bacteria. J Gen Microbiol 61:205–218. https://doi.org/10.1099/00221287-61-2-205.
2. Semrau JD. 2011. Bioremediation via methanotrophy: overview of recent findings and suggestions for future research. Front Microbiol 2:209. https://doi.org/10.3389/fmicb.2011.00209.
3. Stein LY, Yoon S, Semrau JD, DiSpirito AA, Crombie A, Murrell JC, Vuilleumier S, Kalyuzhnaya MG, Op den Camp HJM, Bringel F, Bruce D, Cheng JF, Copeland A, Goodwin L, Han S, Hauser L, Jetten MSM, Lajus A, Land ML, Lapidus A, Lucas S, Médigue C, Pitluck S, Woyke T, Zeytun A, Klotz MG. 2010. Genome sequence of the obligate methanotroph *Methylosinus trichosporium* strain OB3b. J Bacteriol 192:6497–6498. https://doi.org/10.1128/JB.01144-10.
4. Matsen JB, Yang S, Stein LY, Beck D, Kalyuzhnaya MG. 2013. Global molecular analyses of methane metabolism in methanotrophic alphaproteobacterium, *Methylosinus trichosporium* OB3b. Part I: transcriptomic study. Front Microbiol 4:40. https://doi.org/10.3389/fmicb.2013.00040.
5. Yang S, Matsen JB, Konopka M, Green-Saxena A, Clubb J, Sadilek M, Orphan VJ, Beck D, Kalyuzhnaya MG. 2013. Global molecular analyses of methane metabolism in methanotrophic alphaproteobacterium, *Methylosinus trichosporium* OB3b. Part II. Metabolomics and 13C-labeling study. Front Microbiol 4:70. https://doi.org/10.3389/fmicb.2013.00070.
6. 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.
7. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13:238. https://doi.org/10.1186/1471-2105-13-238.
8. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC. 2000. A whole-genome assembly of *Drosophila*. Science 287:2196–2204. https://doi.org/10.1126/science.287.5461.2196.
9. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
10. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
11. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8:64. https://doi.org/10.1186/1471-2105-8-64.
12. Cevallos MA, Cervantes-Rivera R, Gutiérrez-Ríos RM. 2008. The *repABC* plasmid family. Plasmid 60:19–37. https://doi.org/10.1016/j.plasmid.2008.03.001.