Draft Genome Sequences of *Campylobacterales* (*Epsilonproteobacteria*) Obtained from Methanogenic Oil Sands Tailings Pond Metagenomes

BoonFei Tan, Julia Foght

Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

Draft genome sequences of two *Campylobacterales* (*Sulfurospirillum* sp. strain SCADC and *Sulfuricurvum* sp. strain MLSB [Mildred Lake Settling Basin]) were obtained by taxonomic binning of metagenomes originating from an oil sands tailings pond. Both genomes contain *soxABXYZ* genes involved in sulfur oxidation, highlighting their potential roles in sulfur cycling in oil sands tailings ponds.

Environmental sulfur-oxidizing *Epsilonproteobacteria*—specifically, members of the order *Campylobacterales*—are reported to dominate the deep subsurface of a severely biodegraded oil sands reservoir (1). However, their presence and importance have not been documented in the tailings ponds that store wastes from mining the oil sands' ores (2–4).

The metagenome of MLSB (Mildred Lake Settling Basin), an oil sands tailings pond located in northern Alberta, Canada, was obtained by isolating total DNA for high-throughput sequencing using Illumina-HiSeq. Reads were subjected to quality control and sequence assembly using CLC Genomics Workbench version 7 (CLC-Bio, USA). Contigs assembled from the MLSB metagenome and from the metagenome of SCADC (a methanogenic alkane-degrading enrichment culture derived from MLSB [5]) were binned using sequence homology and composition-based methods, including sequence coverage, GC content, and contig length, followed by tetranucleotide frequency and analysis of principal components. Two *Campylobacterales*-associated genomic bins were recovered and subjected to sequence decontamination as previously described (6). Open-reading frames (ORFs) were predicted using Prodigal (7), followed by BLASTx searches against the NCBI NR database and taxon assignment using MEGAN version.5.0 (8). All contigs in both bins had ≥50% of their ORFs affiliated with *Campylobacterales*.

BLASTn comparison of 16S rRNA genes in the SCADC *Campylobacterales* bin indicated that the draft genome is closely (99%) related to *Sulfurospirillum multivorans* DSM 12446 (CP007201). The genome, hereby named *Sulfurospirillum* sp. strain SCADC, is ~2.7 Mbp contained in 38 scaffolds (680–290,000 bp) with average 41% GC content. All 107 single copy genes (ScpG) (9) had best BLAST hits to *Sulfurospirillum*; comparison to *Sulfurospirillum deleyianum* DSM 6946 (2.3 Mbp; CP001816.1; 108 ScpG) and *S. multivorans* DSM 12446 (3.2 Mbp; CP007201.1; 109 ScpG) indicated recovery of a nearly complete genome, even though the reference genomes have a broad size range (2.3 to 3.2 Mbp). Annotation using RAST (10) predicted 2,718 coding DNA sequences (CDSs) in *Sulfurospirillum* SCADC, with 1,695 of 2,718 ORFs

sharing >60% BLASTp similarity with those in *S. deleyianum* DSM 6946, whereas 556 ORFs were present only in *Sulfurospirillum* SCADC.

Genome binning of *Campylobacterales*-affiliated contigs from MLSB did not recover any 16S rRNA genes. However, the draft genome contains 107 of 110 ScpG in *Sulfuricurvum kujiense* DSM 16994 (NC_014762.1), all of which had best BLAST hits to *Sulfuricurvum* spp., indicating recovery of a nearly complete genome. The draft genome (herein named *Sulfuricurvum* sp. MLSB) is ~2.1 Mbp contained in 148 scaffolds (1000–104,000 bp) with average 49% GC content. Of 2,119 predicted CDSs, 1,549 of 2,119 have BLASTp similarity >60% to ORFs in *S. kujiense* DSM 16994, whereas 277 CDSs were present only in *Sulfuricurvum* MLSB.

Sulfurospirillum SCADC and *Sulfuricurvum* MLSB both carry *soxABXYZ* genes for thiosulfate oxidation, plus genes for the Wood-Ljungdahl CO_2 fixation pathway (11). Physiological examination of reference strains shows metabolic potential for using sulfur compounds, nitrate, or H_2 as electron acceptors (12, 13), which are relevant to the biogeochemistry of oil sands tailings ponds (2, 4). Comparative genomics should provide insight into their importance in nutrient cycling and oxidation of reduced sulfur compounds in oil sands tailings ponds.

Nucleotide sequence accession numbers. The whole-genome shotgun projects of *Sulfurospirillum* sp. SCADC and *Sulfuricurvum* sp. MLSB have been deposited at DDBJ/EMBL/GenBank under the accession numbers JQGK00000000 and JQGL00000000, respectively. The versions described in this paper are versions JQGK01000000 and JQGL01000000.

ACKNOWLEDGMENTS

REFERENCES

1. **Hubert CRJ, Oldenburg TBP, Fustic M, Gray ND, Larter SR, Penn K, Rowan AK, Seshadri R, Sherry A, Swainsbury R, Voordouw G, Voordouw JK, Head IM.** 2012. Massive dominance of *Epsilonproteobacteria* in formation waters from a Canadian oil sands reservoir containing severely biodegraded oil. Environ. Microbiol. **14:**387–404. http://dx.doi.org/10.1111/j.1462-2920.2011.02521.x.

2. **Penner TJ, Foght JM.** 2010. Mature fine tailings from oil sands processing harbour diverse methanogenic communities. Can. J. Microbiol. **56:**459–470. http://dx.doi.org/10.1139/W10-029.

3. **An D, Caffrey SM, Soh J, Agrawal A, Brown D, Budwill K, Dong X, Dunfield PF, Foght J, Gieg LM, Hallam SJ, Hanson NW, He Z, Jack TR, Klassen J, Konwar KM, Kuatsjah E, Li C, Larter S, Leopatra V, Nesbø CL, Oldenburg T, Pagé AP, Ramos-Padron E, Rochman FF, Saidi-Mehrabad A, Sensen CW, Sipahimalani P, Song YC, Wilson S, Wolbring G, Wong M-L, Voordouw G.** 2013. Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. Environ. Sci. Technol. **47:**10708–10717. http://dx.doi.org/10.1021/es4020184.

4. **Ramos-Padrón E, Bordenave S, Lin S, Bhaskar IM, Dong X, Sensen CW, Fournier J, Voordouw G, Gieg LM.** 2011. Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond. Environ. Sci. Technol. **45:**439–446. http://dx.doi.org/10.1021/es1028487.

5. **Tan B, Dong X, Sensen CW, Foght J.** 2013. Metagenomic analysis of an anaerobic alkane-degrading microbial culture: potential hydrocarbon-activating pathways and inferred roles of community members. Genome **56:**599–611. http://dx.doi.org/10.1139/gen-2013-0069.

6. **Tan B, Charchuk R, Li C, Nesbo C, Abu-Laban N, Foght J.** Draft genome sequence of uncultivated Firmicutes (Peptococcaceae SCADC) single cells sorted from methanogenic alkane-degrading cultures. Genome Announc., in press.

7. **Hyatt D, Chen G-L, 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.

8. **Huson DH, Auch AF, Qi J, Schuster SC.** 2007. MEGAN analysis of metagenomic data. Genome Res. **17:**377–386. http://dx.doi.org/10.1101/gr.5969107.

9. **Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH.** 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. **31:**533–538. http://dx.doi.org/10.1038/nbt.2579.

10. **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. http://dx.doi.org/10.1093/nar/gkt1226.

11. **Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahrén D, Tsoka S, Darzentas N, Kunin V, López-Bigas N.** 2005. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res. **33:**6083–6089. http://dx.doi.org/10.1093/nar/gki892.

12. **Han C, Kotsyurbenko O, Chertkov O, Held B, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng J-F, Tapia R, Goodwin LA, Pitluck S, Liolios K, Pagani I, Ivanova N, Mavromatis K, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang Y-J, Jeffries CD, Brambilla E-M, Rohde M, Spring S, Sikorski J, Göker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk H-P, Detter JC.** 2012. Complete genome sequence of the sulfur compounds oxidizing chemolithoautotroph *Sulfuricurvum kujiense* type strain (YK-1^T). Stand Genomics Sci. **6:**94–103. http://dx.doi.org/10.4056/sigs.2456004.

13. **Sikorski J, Lapidus A, Copeland A, Glavina Del Rio T, Nolan M, Lucas S, Chen F, Tice H, Cheng J-F, Saunders E, Bruce D, Goodwin L, Pitluck S, Ovchinnikova G, Pati A, Ivanova N, Mavromatis K, Chen A, Palaniappan K, Chain P, Land M, Hauser L, Chang Y-J, Jeffries CD, Brettin T, Detter JC, Han C, Rohde M, Lang E, Spring S, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk H-P.** 2010. Complete genome sequence of *Sulfurospirillum deleyianum* type strain (5175^T). Stand Genomics Sci. **2:**149–157. http://dx.doi.org/10.4056/sigs.671209.