First Draft Genome Sequence from a Member of the Genus Agrococcus, Isolated from Modern Microbialites

Richard Allen White III,a Christopher J. Grassa,b Curtis A. Suttleab,c,d

Department of Microbiology & Immunology, University of British Columbia, Vancouver, British Columbia, Canada; Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada; Department of Earth, Ocean & Atmospheric Science, University of British Columbia, Vancouver, British Columbia, Canada; Canadian Institute for Advanced Research

We report the first draft genome sequence from a member of the genus Agrococcus, isolated from cold thrombolytic microbialites within Pavilion Lake, British Columbia, Canada. The draft genome assembly for Agrococcus pavilionensis strain RW-1 has a size of 2,878,403 bp with a G+C content of 72.56%.

Received 7 May 2013 Accepted 21 May 2013 Published 27 June 2013
Copyright © 2013 White et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.
Address correspondence to Curtis A. Suttle, suttle@science.ubc.ca.

Agrococcus pavilionensis strain RW-1 is the olive-colored (3.3 μg liter⁻¹ total phosphorus) Pavilion Lake (50.8677°N, 121.7419°W), which lies in Marble Canyon near Lillooet, British Columbia, Canada. Pavilion Lake harbors a diverse array of modern microbialites, which are contemporary biogenically derived carbonate structures (10, 11). Pavilion Lake microbialites consist mainly of clotted and nonlayered thrombolic structures (10) that occur in the permanently cold (4 to 8°C) water deeper than 5 m (11). A. pavilionensis was isolated from a cabbage-shaped thrombolite collected at a depth of 20 m. DNA was extracted using QiaGen QIAamp followed by QiaGen MinElute cleanup columns. The Illumina MiSeq library was constructed using the Lucigen NxSeq library prep kit without final PCR enrichment.

Whole-genome shotgun sequencing was completed using Illumina MiSeq in the 250-bp paired-read format. A partial flow cell output was 2.89 million raw reads with 713,936,519 bp of raw sequence. Paired reads were error corrected and connected using AllPaths-LG (version 44837) (12). In the data set, 31mers were counted using Jellyfish (version 1.1.10) (13). Reads that contained 31mers with a multiplicity of >1,250 were partitioned for de novo assembly. The partitioned reads were assembled using Celera assembler 7.0 (14). The high-copy reads assembled as a single contig of 1,427 bp in length representing a high-copy plasmid. The remaining reads assembled as 50 contigs summing 2,878,403 bp (N₅₀ length, 133,224; N₉₀ length, 31,609; G+C content, 72.56%). The 16S rRNA gene sequence was confirmed by Sanger sequencing and was found to have 99.99% identity to the 16S rRNA gene predicted from the draft genome.

Annotation was conducted on the RAST server using the Glimmer 3 option (15) and predicted 2,506 protein-coding genes, including 48 noncoding RNA genes and 126 predicted SEED subsystem features. The potential to metabolize a wide range of carbon compounds is predicted from the genome, including D-ribose, fructose, lactate, glycerate, chitin, deoxyribose, and deoxyribonucleoside catabolism. Genes related to those encoding the phosphate (Pho) regulon for high-affinity uptake of phosphate and cold shock proteins were also found.

Further analysis of the genome, including functional and biochemical measurements, will be used to understand the possible roles of A. pavilionensis in the highly diverse microbial community contained within the Pavilion Lake microbialites. This is the first draft genome for the genus Agrococcus, which will provide a template for many further phylogenetic, comparative genomic, metagenomic, and functional studies of this widely distributed genus.

Nucleotide sequence accession numbers. This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ASHR01000000. The version described in this paper is version ASHR01000000.

ACKNOWLEDGMENTS

We thank Sugandha Dandekar (Uma) and Hemani Wijesuriya at the UCLA Sequencing & Genotyping Core for the excellent sequencing data and Donnie Reid and the divers and boat operators from the Pavilion Lake Research Project who collected the sample from which the bacterial strain was isolated.

Financial support was provided by the MARSLife Project (9F052-10-
REFERENCES


