Genome sequence of strain HIMB30, a novel member of the marine Gammaproteobacteria

Megan Huggett
Michael Rappe
Genome Sequence of Strain HIMB30, a Novel Member of the Marine Gammaproteobacteria

Megan J. Huggett and Michael S. Rappé
Hawaii Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Kaneohe, Hawaii, USA

Strain HIMB30 was isolated from coastal Hawaii seawater by extinction culturing in seawater-based oligotrophic medium. It is a phylogenetically unique member of the class Gammaproteobacteria that is only distantly related to its closest cultured relatives. Here we present the genome sequence of strain HIMB30, including genes for proteorhodopsin-based phototrophy and the Calvin-Benson-Bassham cycle.

The genome of HIMB30 was shotgun sequenced by the J. Craig Venter Institute as part of the Moore Foundation Microbial Genome Sequencing Project (http://camera.calit2.net/microgenome), resulting in 629,573 reads of 454 FLX Titanium sequence data. A draft, unclosed genome was assembled to 102\% coverage from 624,954 reads (254,771,663 bases) of sequence data. Functional annotation was performed with the Integrated Microbial Genomes and Expert Review (IMG-ER) pipeline (9). Genes identified using Prodigal (5) were translated and used to search the NCBI nonredundant protein database and the UniProt, TIGRFam, and Rfam (4) databases. Non-annotated genes were predicted using tRNAscanSE (8), Rfammer (7), and Pfam (4). Additional gene prediction analysis and manual functional annotation were performed within IMG-ER.

The draft genome of HIMB30 consists of 36 contigs ranging in size from 266 to 638,152 bases, resulting in a total genome of 2,168,870 bases. The G+C content is 49.86\%. In all, 2,295 predicted open reading frames and 2,249 predicted protein-coding genes were identified, of which 1,874 have a predicted function. There are predicted single copies of the 5S, 16S, and 23S rRNA genes and 36 predicted tRNAs. Genes for a complete tricarboxylic acid cycle are present in the genome of HIMB30, but it possesses lesions in both the Embden-Meyerhoff-Parnas pathway and the oxidative portion of the pentose phosphate pathway. Genes for two key enzymes of the Entner-Doudoroff pathway, phosphogluconate dehydratase and 2-keto-3-deoxyphosphogluconate aldolase, were found. The genome contains putative genes for the synthesis of all of the essential amino acids and a number of vitamins, including para-aminobenzoic acid, lipoic acid, pyridoxine (vitamin B6), riboflavin, pantothenic acid, and coenzyme A. In addition to genes for carbon monoxide and sulfur oxidation, genes for proteorhodopsin-based phototrophy and CO2 fixation via the Calvin-Benson-Bassham cycle were found.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited in GenBank under accession no. AGIG00000000. The version described in this paper is the first version, AGIG01000000. The genome project was also deposited in the Genomes OnLine Database as project Gb03571, and the NCBI taxonomy identification number is 751994.

ACKNOWLEDGEMENTS

We thank Steve Ferriera, Justin Johnson, and other J. Craig Venter Institute scientists for performing the sequencing and assembly of this genome, with support provided by the Gordon and Betty Moore Foundation Microbial Genome Sequencing Project (http://www.moore.org/marinemicro). Strain preparation, genomic DNA extraction, and annotation were supported by the Center for Microbial Oceanography: Research and Education (NSF Science and Technology Center award EF-0424599).

REFERENCES

8. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detect-

