Genome Sequence of *Halomonas* sp. Strain MCTG39a, a Hydrocarbon-Degrading and Exopolymeric Substance-Producing Bacterium

Tony Gutierrez,[a] William B. Whitman,[b] Marcel Huntemann,[c] Alex Copeland,[c] Amy Chen,[c] Nikos Kyrpides,[c] Victor Markowitz,[c] Manoj Pillay,[c] Natalia Ivanova,[c] Natalia Mikhailova,[c] Galina Ovchinnikova,[c] Evan Andersen,[c] Amrita Pati,[c] Dimitrios Stamatis,[c] T. B. K. Reddy,[c] Chew Yee Ngan,[c] Mansi Chovatia,[c] Chris Daum,[c] Nicole Shapiro,[c] Michael N. Cantor,[c] Tanja Woyke[c]

School of Life Sciences, Heriot-Watt University, Edinburgh, United Kingdom[a]; Department of Microbiology, University of Georgia, Athens, Georgia, USA[b]; DOE Joint Genome Institute, Walnut Creek, California, USA[c]

Halomonas sp. strain MCTG39a was isolated from coastal sea surface water based on its ability to utilize *n*-hexadecane. During growth in marine medium the strain produces an amphiphilic exopolymeric substance (EPS) amended with glucose, which emulsifies a variety of oil hydrocarbon substrates. Here, we present the genome sequence of this strain, which is 4,979,193 bp with 4,614 genes and an average G+C content of 55.0%.

Address correspondence to Tony Gutierrez, tony.gutierrez@hw.ac.uk.

*H*alamonas sp. strain MCTG39a from a sea surface water sample collected off the coast of California was isolated by enrichment with *n*-hexadecane as the sole carbon source (K. Salek, E. Sheehan, and T. Gutierrez, unpublished results). The strain produces a surface-active (amphiphilic) exopolymer, which emulsifies a range of hydrocarbons and increases their bioavailability for biodegradation. Strain MCTG39a is a strictly aerobic and motile, rod-shaped marine bacterium that degrades hydrocarbons. It produces an exopolymeric substance (EPS) with strong emulsifying qualities against a range of hydrocarbons.

Here, we report the genome sequence of *Halomonas* sp. strain MCTG39a. Genomic DNA was isolated, and the sequence was generated at the Department of Energy (DOE) Joint Genome Institute (JGI) using Pacific Biosciences (PacBio) technology. A Pacbio SMRTbellTM library was constructed and sequenced on the PacBio RS platform, which generated 247,476 filtered subreads totaling 781.1 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The raw reads were assembled using HGAP (version: 2.1.1) (1). The final draft assembly produced 2 scaffolds containing 2 contigs a total of 5.0 Mbp in size with input read coverage of 174.6×.

Project information is available in the Genomes OnLine Database (2). Genes were identified using Prodigal (3), followed by a round of manual curation using GenePRIMP (4) as part of the JGI's microbial annotation pipeline (5). The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database and the UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (6) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes from SILVA (7). Other noncoding RNAs, such as the RNA components of the protein secretion complex and the RNase P, were identified by searching the genome for the corresponding Rfam profiles using Infernal (http://infernal.janelia.org). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes–Expert Review (IMG ER) platform (http://img.jgi.doe.gov) developed by the Joint Genome Institute, Walnut Creek, CA, USA (8).

The complete genome sequence length is 4,979,193 bp with a G+C content of 55.0%. The genome contains 4,614 genes (4,522 protein-coding genes) with functional predictions for 3,957 of them. A total of 92 RNA genes were detected. Other genes, characteristic for the genus, are given in the IMG database (8). This genome sequence is expected to provide great insights into the unusual life style of this organism.

Nucleotide sequence accession numbers. The draft genome sequence of *Halomonas* sp. strain MCTG39a obtained in this study was deposited in GenBank as part of BioProject no. PRJNA224116, with individual genome sequences submitted as whole-genome shotgun projects under the accession no. JQLV00000000.

REFERENCES

1. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.

2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563–569. http://dx.doi.org/10.1038/nmeth.2474.

2. **Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC.** 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res **43:** D1099–D1106. http://dx.doi.org/10.1093/nar/gku950.

3. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**. http://dx.doi.org/10.1186/1471-2105-11-119.

4. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods **7:**455–457. http://dx.doi.org/10.1038/nmeth.1457.

5. **Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC.** 2009. Standard operating procedure for the annotations of microbial genomes by the Production Genomic Facility of the DOE JGI. Stand Genomic Sci **1:**63–67. http://dx.doi.org/10.4056/sigs.632.

6. **Lowe TM, Eddy SR.** 1997. tRNAscan SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955–964. http://dx.doi.org/10.1093/nar/25.5.0955.

7. **Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J, Glöckner FO.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35:**2188–7196. http://dx.doi.org/10.1093/nar/gkm864.

8. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271–2278. http://dx.doi.org/10.1093/bioinformatics/btp393.