UC Merced

UC Merced Previously Published Works

Authors

Gutierrez, Tony

Whitman, William B

Huntemann, Marcel

et al.

Genome Sequence of *Marinobacter* sp. Strain MCTG268 Isolated from the Cosmopolitan Marine Diatom *Skeletonema costatum*

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]

***Marinobacter* sp. strain MCTG268 was isolated from the cosmopolitan marine diatom *Skeletonema costatum* and can degrade oil hydrocarbons as sole sources of carbon and energy. Here, we present the genome sequence of this strain, which is 4,449,396 bp with 4,157 genes and an average G+C content of 57.0%.**

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

*M*arinobacter sp. strain MCTG268 was isolated from a laboratory culture of the marine diatom *Skeletonema costatum* (CCAP 1077/1C) by enrichment with naphthalene as the sole carbon source. Based on 16S rRNA gene sequence identity, the closest type species is *Marinobacter algicola* strain DG893, which had also been isolated from a laboratory culture of eukaryotic phytoplankton (1).

Here, we report the genome sequence of *Marinobacter* sp. strain MCTG268. Genomic DNA was sequenced through the DOE Joint Genome Institute 2014 Genomic Encyclopedia of Type Strains, Phase III study (2) using Pacific Biosciences (PacBio) technology. A Pacbio SMRTbellTM library was constructed and sequenced on the PacBio RS platform, which generated 220,290 filtered subreads totaling 696.2 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) (3). The final draft assembly produced 14 scaffolds containing 14 contigs totaling 4.4 Mbp in size and input read coverage of 222.5×.

Project information is available in the Genomes Online Database (4). Genes were identified using Prodigal (5), followed by a round of manual curation using GenePRIMP (6) as part of JGI's microbial annotation pipeline (7). The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (8) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA (9). 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 (10).

The complete genome sequence length was 4,449,396 bp with a G+C content of 57.0%. The genome contains 4,157 genes (4,083 protein-coding genes) with functional predictions for 3,388 of them. A total of 74 RNA genes were detected. Other genes, characteristic for the genus, are given in the IMG database (10).

Accession number(s). The draft genome sequence of *Marinobacter* sp. strain MCTG268 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. JQMK00000000.

REFERENCES

1. **Green DH, Bowman JP, Smith EA, Gutierrez T, Bolch CJ.** 2006. *Marinobacter algicola* sp. nov., isolated from laboratory cultures of paralytic shellfish toxin-producing dinoflagellates. Int J Syst Evol Microbiol **56:** 523–527. http://dx.doi.org/10.1099/ijs.0.63447-0.

2. **Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, De Vos P, Vandamme P, Eisen JA, Garrity G, Hugenholtz P, Kyrpides NC.** 2015. Genomic encyclopedia of bacterial and archaeal type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains. Stand Genomic Sci **10:**26. http://dx.doi.org/10.1186/s40793-015-0017-x.

3. **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.

4. **Reddy TBK, 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.

5. **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:**119. http://dx.doi.org/10.1186/1471-2105-11-119.

6. **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.

7. **Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen IM-A, Pati A, Nielsen T, Markowitz VM, Kyrpides NC.** 2015. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). Stand Genom Sci **10:**86. http://dx.doi.org/10.1186/s40793-015-0077-y.

8. **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.

9. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glckner 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.

10. **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.