



Genome Sequence of *Oceanicola* sp. Strain MCTG156(1a), Isolated from a Scottish Coastal Phytoplankton Net Sample

Tony Gutierrez,^a William B. Whitman,^b Marcel Huntemann,^c Alex Copeland,^c Amy Chen,^c Neha Vargese,^c Nikos C. Kyrpides,^c Manoj Pillay,^c Natalia Ivanova,^c Natalia Mikhailova,^c Supratim Mukherjee,^c Dimitrios Stamatis,^c T. B. K. Reddy,^c Chew Yee Ngan,^c Mansi Chovatia,^c Chris Daum,^c Nicole Shapiro,^c Tanja Woyke^c

School of Engineering and Physical 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

ABSTRACT *Oceanicola* sp. strain MCTG156(1a) was isolated from a phytoplankton net sample collected on the west coast of Scotland and selected based on its ability to degrade polycyclic aromatic hydrocarbons. Here, we present the genome sequence of this strain, which comprises 3,881,122 bp with 3,949 genes and an average G+C content of 62.7%.

Oceanicola sp. strain MCTG156(1a) was isolated from a phytoplankton net sample that was trawled in 2009 at a sampling station designated LY1, located on the west coast of Scotland near Oban, Argyll. The strain was isolated by enrichment with phenanthrene in Zobell's 2216 marine medium at 10-fold dilution. Colonies on agar plates sprayed with phenanthrene produced distinct halos that indicated the strain's ability to degrade the hydrocarbon. Based on 16S rRNA gene sequence identity, the closest type species was *Oceanicola pacificus* strain W11-2B^T, which had been isolated from a pyrene-degrading consortium that was enriched from sediment from the Pacific Ocean (1).

Here, we report the genome sequence of *Oceanicola* sp. strain MCTG156(1a). Genomic DNA was sequenced through the DOE Joint Genome Institute (JGI) 2014 Genomic Encyclopedia of Type Strains, Phase III study (2), using the Pacific Biosciences (PacBio) technology. A PacBio SMRTbell library was constructed and sequenced on the PacBio RS platform, which generated 239,103 filtered subreads totaling 750.9 Mb. 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 5 scaffolds containing 5 contigs totaling 3.9 Mb in size with an input read coverage of 217.9×.

Project information is available in the Genomes OnLine Database (4). Genes were identified using Prodigal (5), as part of the JGI's microbial annotation pipeline (6). The predicted coding sequences (CDSs) were translated and used to search the NCBI nonredundant, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (7) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA (8). Other noncoding RNAs, such as the RNA components of the protein secretion complex and RNase P, were identified by searching the genome for the corresponding Rfam profiles using Infernal (<http://infernal.janelia.org>). Additional analysis and manual functional annotation was performed within the Integrated Microbial Genomes—Expert Review (IMG ER) platform (<https://img.jgi.doe.gov>) developed by the JGI (Walnut Creek, CA, USA) (9).

Received 28 June 2017 Accepted 3 July 2017 Published 10 August 2017

Citation Gutierrez T, Whitman WB, Huntemann M, Copeland A, Chen A, Vargese N, Kyrpides NC, Pillay M, Ivanova N, Mikhailova N, Mukherjee S, Stamatis D, Reddy TBK, Ngan CY, Chovatia M, Daum C, Shapiro N, Woyke T. 2017. Genome sequence of *Oceanicola* sp. strain MCTG156(1a), isolated from a Scottish coastal phytoplankton net sample. *Genome Announc* 5:e00796-17. <https://doi.org/10.1128/genomeA.00796-17>.

Copyright © 2017 Gutierrez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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

The complete genome sequence length was 3,881,122 bp with a G+C content of 62.7%. The genome contained 3,949 genes (3,881 protein-coding genes) with functional predictions for 3,226 of them. A total of 68 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 *Oceanicola* sp. strain MCTG156(1a) 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. [JQMY00000000](https://doi.org/10.1099/jqs.0.003400-0).

ACKNOWLEDGMENTS

T.G. was supported by a Marie Curie International Outgoing Fellowship (PIOF-GA-2008-220129) within the 7th European Community Framework Program.

This work was conducted by the U.S. Department of Energy, Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

REFERENCES

1. Yuan J, Lai Q, Wang B, Sun F, Liu X, Du Y, Li G, Gu L, Zheng T, Shao Z. 2009. *Oceanicolapacificus* sp. nov., isolated from a deep-sea pyrene-degrading consortium. *Int J Syst Evol Microbiol* 59:1158–1161. <https://doi.org/10.1099/jqs.0.003400-0>.
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. <https://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. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
4. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezhemskaya O, Isbandi M, Thomas AD, Ali R, Sharma K, Kyrpides NC, Reddy TB. 2017. Genomes online database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res* 45:D446–DD456. <https://doi.org/10.1093/nar/gkw992>.
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. <https://doi.org/10.1186/1471-2105-11-119>.
6. 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. <https://doi.org/10.1186/s40793-015-0077-y>.
7. 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.
8. Pruesse E, Quast C, Knittel FB, 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* 35:7188–7196. <https://doi.org/10.1093/nar/gkm864>.
9. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25:2271–2278. <https://doi.org/10.1093/bioinformatics/btp393>.
10. Chen IA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A, Huang J, Andersen E, Huntemann M, Varghese N, Hadjithomas M, Tennesen K, Nielsen T, Ivanova NN, Kyrpides NC. 2017. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* 45:D507–DD516. <https://doi.org/10.1093/nar/gkw929>.