

1 **Genome sequence of *Oceanicola* sp. strain MCTG156(1a) isolated**  
2 **from a Scottish coastal phytoplankton net sample**

3

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

8

9 <sup>a</sup> School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh,  
10 UK;

11 <sup>b</sup> Department of Microbiology, University of Georgia, Athens, Georgia, USA;

12 <sup>c</sup> DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698,  
13 USA.

14

15 **\*Correspondence:**

16 Dr. Tony Gutierrez

17 School of Engineering and Physical Sciences

18 Heriot-Watt University

19 Edinburgh EH14 4AS, U.K.

20 [tony.gutierrez@hw.ac.uk](mailto:tony.gutierrez@hw.ac.uk)

21

22 **Running title:** Genome sequence of a *Oceanicola* species

23

24

25

26 **Abstract**

27 *Oceanicola* sp. strain MCTG156(1a) was isolated from a phytoplankton net sample  
28 collected on the west coast of Scotland and selected based on its ability to degrade  
29 hydrocarbons. Here, we present the genome sequence of this strain, which is  
30 3,881,122 bp with 3,949 genes and an average G+C content of 62.7%.

31

32

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

42 Here, we report the genome sequence of *Oceanicola* sp. strain MCTG156(1a).  
43 Genomic DNA was sequenced through the DOE Joint Genome Institute 2014  
44 Genomic Encyclopedia of Type Strains, Phase III study (2) using the Pacific  
45 Biosciences (PacBio) technology. A Pacbio SMRTbell<sup>TM</sup> library was constructed  
46 and sequenced on the PacBio RS platform, which generated 239,103 filtered subreads  
47 totaling 750.9 Mbp. All general aspects of library construction and sequencing  
48 performed at the JGI can be found at <http://www.jgi.doe.gov>. The raw reads were  
49 assembled using HGAP (version: 2.1.1) (3). The final draft assembly produced 5

50 scaffolds containing 5 contigs totaling 3.9 Mbp in size and input read coverage of  
51 217.9X.

52 Project information is available in the Genomes OnLine Database (4). Genes  
53 were identified using Prodigal (5), as part of the JGI's microbial annotation pipeline  
54 (6). The predicted coding sequences (CDSs) were translated and used to search the  
55 National Center for Biotechnology Information (NCBI) nonredundant database,  
56 UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE  
57 tool (7) was used to find tRNA genes, whereas ribosomal RNA genes were found by  
58 searches against models of the ribosomal RNA genes built from SILVA (8). Other  
59 noncoding RNAs, such as the RNA components of the protein secretion complex and  
60 RNase P, were identified by searching the genome for the corresponding Rfam  
61 profiles using INFERNAL (<http://inferral.janelia.org>). Additional analysis and  
62 manual functional annotation was performed within the Integrated Microbial  
63 Genomes–Expert Review (IMG ER) platform (<http://img.jpi.doe.gov>) developed by  
64 the Joint Genome Institute, Walnut Creek, CA, USA (9).

65 The complete genome sequence length was 3,881,122 bp with a G+C content  
66 of 62.7%. The genome contained 3,949 genes (3,881 protein-coding genes) with  
67 functional predictions for 3,226 of them. A total of 68 RNA genes were detected.  
68 Other genes, characteristic for the genus, are given in the IMG database (10).

69 **Nucleotide sequence accession number.** The draft genome sequence of  
70 *Oceanicola* sp. strain MCTG156(1a) obtained in this study was deposited in GenBank  
71 as part of BioProject no. PRJNA224116, with individual genome sequences submitted  
72 as whole-genome shotgun projects under the accession no. JQMY00000000.

73

74

75 **ACKNOWLEDGMENT**

76 T.G. was supported by a Marie Curie International Outgoing Fellowship (PIOF-GA-  
77 2008-220129) within the 7th European Community Framework Program. Partial  
78 support was also provided from the European Union's Horizon 2020 research and  
79 innovation programme under grant agreement No. 635340.

Commented [TG1]: I need to add this to the final version when accepted.

80 The work was conducted by the U.S. Department of Energy Joint Genome  
81 Institute, a DOE Office of Science User Facility, which is supported by the Office of  
82 Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

83

84 **REFERENCES**

- 85 1. Yuan J, Lai Q, Wang B, Sun F, Liu X, Du Y, Li G, Gu L, Zheng T, Shao Z.  
86 2009. *Oceanicola pacificus* sp. nov., isolated from a deep-sea pyrene-degrading  
87 consortium. *Int J Syst Evol Microbiol* 59:1158-1161.
- 88 2. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, De Vos P,  
89 Vandamme P, Eisen JA, Garrity G, Hugenholtz P, Kyrpides NC. 2015.  
90 Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the  
91 genomes of soil and plant-associated and newly described type strains.  
92 *Standards Genomic Sci* 10:26.
- 93 3. Chin C, Alexander D, Marks P, Klammer A, Drake J, Heiner C, Clum A,  
94 Copeland A, Huddleston J, Eichler E, Turner S, Korlach J. 2013. Nonhybrid,  
95 finished microbial genome assemblies from long-read SMRT sequencing data.  
96 *Nat Methods* 10:563-569.
- 97 4. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemskaja O, Isbandi  
98 M, Thomas A, Ali RS, Kyrpides NC, and Reddy TBK. 2017. *Genomes OnLine*

99 Database (GOLD) v.6: Data updates and feature enhancements. *Nucleic Acids*  
100 *Res.* 45(D1):D446-D456.

101 5. Hyatt D, Chen GL, Lacascio PF, Land ML, Larimer FW, Hauser LJ. 2010.  
102 Prodigal: prokaryotic gene recognition and translation initiation site  
103 identification. *BMC Bioinformatics* 11:119.

104 6. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D,  
105 Palaniappan K, Szeto E, Pillay M, Chen IM-A, Pati A, Nielsen T, Markowitz  
106 VM, Kyrpides NC. 2015. The Standard Operating Procedure of the DOE-JGI  
107 Microbial Genome Annotation Pipeline (MGAP v.4) *Stand Genom Sci.* 2015,  
108 10:86.

109 7. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of  
110 transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955-964.

111 8. Pruesse E, Quast C, Knittel, Fuchs B, Ludwig W, Peplies J, Glckner FO. 2007.  
112 SILVA: a comprehensive online resource for quality checked and aligned  
113 ribosomal RNA sequence data compatible with ARB. *Nuc Acids Res* 35: 2188-  
114 7196.

115 9. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC.  
116 2009. IMG ER: a system for microbial genome annotation expert review and  
117 curation. *Bioinformatics* 25:2271-2278.

118 10. Chen IA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A,  
119 Huang J, Andersen E, Huntemann M, Varghese N, Hadjithomas M, Tennesen  
120 K, Nielsen T, Ivanova NN, Kyrpides NC. 2017. IMG/M: integrated genome and  
121 metagenome comparative data analysis system. *Nucleic Acids Res.*  
122 45(D1):D507-D516.