Genome Sequence of Roseovarius sp. Strain MCTG156(2b) Isolated from a Phytoplankton Net Trawl on the Scottish West Coast

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

Published in:
Genome Announcements

DOI:
10.1128/genomeA.00837-17

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Heriot-Watt University Research Portal

Citation for published version (APA):
Gutierrez, T., Whitman, W. B., Huntemann, M., Copeland, A., Chen, A., Vargese, N., ... Woyke, T. (2017). Genome Sequence of Roseovarius sp. Strain MCTG156(2b) Isolated from a Phytoplankton Net Trawl on the Scottish West Coast. Genome Announcements, 5(35), [e00837-17]. DOI: 10.1128/genomeA.00837-17
Genome Sequence of *Roseovarius* sp. Strain MCTG156(2b) Isolated from a Phytoplankton Net Trawl on the Scottish West Coast

Tony Gutierrez,a William B. Whitman,b Marcel Huntemann,c Alex Copeland,c Amy Chen,a Neha Vargese,c Nikos C. Kyprides,c Manoj Pillay,c Natalia Ivanova,c Natalia Mikhailova,a 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 Woykec

School of Engineering and Physical Sciences, Heriot Watt University, Edinburgh, United Kingdoma; Department of Microbiology, University of Georgia, Athens, Georgia, USAb; DOE Joint Genome Institute, Walnut Creek, California, USAc

ABSTRACT *Roseovarius* sp. strain MCTG156(2b) was isolated from a phytoplankton net sample collected on the west coast of Scotland and was selected based on its ability to degrade polycyclic aromatic hydrocarbons. Here, we present the genome sequence of this strain, which is 5,113,782 bp, with 5,142 genes and an average G+C content of 60.7%.

*R*oseovarius sp. strain MCTG156(2b) 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 a 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 *Roseovarius mucosus* strain DFL-24, which had also been isolated from a laboratory culture of a dinoflagellate (1).

Here, we report the genome sequence of *Roseovarius* sp. strain MCTG156(2b). Genomic DNA was sequenced through the DOE Joint Genome Institute 2014 Genomic Encyclopaedia of Type Strains phase III study (2) using Pacific Biosciences (PacBio) technology. A PacBio SMRTbell library was constructed and sequenced on the PacBio RS platform, which generated 170,293 filtered subreads totaling 598.2 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 8 scaffolds containing 8 contigs totaling 5.1 Mbp in size, with input read coverage of 121.5X.

Project information is available in the Genomes OnLine Database (4). Genes were identified using Prodigal (5), as part of the Joint Genome Institute (JGI) Microbial Annotation Pipeline (6). 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 tRNAscan-SE tool (7) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA 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 Pfam profiles using Infernal (http://infernal.janelia.org). Additional analysis and manual functional annotation were performed within the Integrated Microbial Genomes-Expert Review Program.
(IMG-ER) platform (https://img.jgi.doe.gov/) developed by the Joint Genome Institute, Walnut Creek, CA, USA (9).

The complete genome sequence length was 5,113,782 bp, with a G+C content of 60.7%. The genome contained 5,142 genes (5,078 protein-coding genes), with functional predictions for 4,198 of them. A total of 64 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 *Roseovarius* sp. strain MCTG156(2b) 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. JQLS00000000.

**ACKNOWLEDGMENTS**

T.G. was supported by a Marie Curie International Outgoing Fellowship (PIOF-GA-2008-220129) within the Seventh European Community Framework Program. Partial support was also provided by the European Union’s Horizon 2020 research and innovation program under grant agreement no. 635340. The 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**