UC Merced
UC Merced Previously Published Works

Title

Genome Sequence of Arenibacter algicola Strain TG409, a Hydrocarbon-Degrading Bacterium Associated with Marine Eukaryotic Phytoplankton

Permalink

Journal

Microbiology Resource Announcements, 4(4)

ISSN

2576-098X

Authors

Gutierrez, Tony
Whitman, William B
Huntemann, Marcel
et al.

Publication Date

2016-08-25

DOI

10.1128/genomea.00765-16

Peer reviewed

Genome Sequence of *Arenibacter algicola* Strain TG409, a Hydrocarbon-Degrading Bacterium Associated with Marine Eukaryotic Phytoplankton

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]

Arenibacter algicola strain TG409 was isolated from *Skeletonema costatum* and exhibits the ability to utilize polycyclic aromatic hydrocarbons as sole sources of carbon and energy. Here, we present the genome sequence of this strain, which is 5,550,230 bp with 4,722 genes and an average G+C content of 39.7%.

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

*A*renibacter algicola strain TG409 was isolated from a laboratory culture of the marine diatom *Skeletonema costatum* (CCAP 1077/1C) by enrichment with polycyclic aromatic hydrocarbons (PAHs) as the sole carbon source (1). The strain represents the first *Arenibacter* species associated with a laboratory culture of *S. costatum* and shows the ability to degrade PAHs as a sole source of carbon and energy (1). Notably, work with this and other *Arenibacter* type strains has shown that PAH degradation may be a shared phenotype among members of this genus (1). Strain TG409 is a strictly aerobic and nonmotile rod-shaped bacterium that is associated with the cosmopolitan marine diatom *S. costatum*. It produces amphiphilic biopolymers (e.g., biosurfactants) on the cell surface that may facilitate attachment to oil droplets and increase the bioavailability of hydrocarbons (1).

Here, we report the genome sequence of *Arenibacter algicola* strain TG409. Genomic DNA was isolated, and the sequence generated at the Department of Energy (DOE) Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) technology. A Pacbio SMRTbell library was constructed and sequenced on the PacBio RS platform, which generated 292,099 filtered subreads totaling 1074.1 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) (2). The final draft assembly produced three scaffolds containing three contigs totaling 5.6 Mbp in size and input read coverage of 185.2×.

Project information is available in the Genomes Online Database (3). Genes were identified using Prodigal (4), followed by a round of manual curation using GenePRIMP (5) as part of the JGI's microbial annotation pipeline (6). 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 (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 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 (https://img.jgi.doe.gov/) developed by the Joint Genome Institute, Walnut Creek, CA, USA (9).

The complete genome sequence length was 5,550,230 bp with a G+C content of 39.7%. The genome contains 4,722 genes (4,649 protein-coding genes) with functional predictions for 3,538 of them. A total of 73 RNA genes were detected. Other genes, characteristic for the genus, are given in the IMG database (9).

Nucleotide sequence accession number. The draft genome sequence of *A. algicola* strain TG409 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. JPOO00000000.

FUNDING INFORMATION

REFERENCES

1. **Gutierrez T, Rhodes G, Mishamandani S, Berry D, Whitman WB, Nichols PD, Semple KT, Aitken MD**. 2014. Polycyclic aromatic hydrocarbon degradation of phytoplankton-associated *Arenibacter* spp. and description of *Arenibacter algicola* sp. nov., an aromatic hydrocarbon-degrading bacterium. Appl Environ Microbiol **80**:618–628. http://dx.doi.org/10.1128/AEM.03104-13.

2. **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.

3. **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.

4. **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.

5. **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.

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. http://dx.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. http://dx.doi.org/10.1093/nar/25.5.0955.

8. **Pruesse E, Quast C, Knittel, Fuchs B, 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.

9. **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.