

Draft Genome Sequence of *Arthrobacter enclensis* NCIM 5488^T for Secondary Metabolism

Priya S. Neurgaonkar, Mahesh S. Dharne, Syed G. Dastager

National Collection of Industrial Microorganisms, CSIR-National Chemical Laboratory, Pune, Maharashtra, India

Here, we report the draft genome sequence of *Arthrobacter enclensis* NCIM 5488^T, an actinobacterium isolated from a marine sediment sample from Chorao Island, Goa, India. This draft genome sequence consists of 4,226,231 bp with a G+C content of 67.08%, 3,888 protein-coding genes, 50 tRNAs, and 10 rRNAs. Analysis of the genome using bioinformatics tools such as antiSMASH and NaPDoS showed the presence of many unique natural product biosynthetic gene clusters.

Received 20 April 2016 Accepted 21 April 2016 Published 2 June 2016

Citation Neurgaonkar PS, Dharne MS, Dastager SG. 2016. Draft genome sequence of *Arthrobacter enclensis* NCIM 5488^T for secondary metabolism. *Genome Announc* 4(3): e00497-16. doi:10.1128/genomeA.00497-16.

Copyright © 2016 Neurgaonkar 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 Syed G. Dastager, sg.dastager@ncl.res.in.

The genus *Arthrobacter* was established by Conn and Dimmick (1) and includes most of the bacteria that exhibit a rod (in young cultures)–coccus (in older cultures) morphological cycle, although some members of the genus are spheres, occurring in pairs and tetrads (2). The unique adaptation characteristic of actinomycetes in the marine environment is a source of interesting research for new species and a promising source of pharmaceutically important compounds (3). The type species *Arthrobacter enclensis* NCIM 5488^T is a Gram-positive aerobic cocci-rod actinobacterium, isolated from a marine sediment sample from Chorao Island, Goa, India (4).

The genomic DNA of the isolates was extracted from 24-h-old tryptone soy agar cultures. The draft genome of *Arthrobacter enclensis* NCIM 5488^T was generated at the DOE Joint Genome Institute (JGI), Walnut Creek, California, USA, using Illumina technology (5). An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform, which generated 8,861,546 reads totaling 1,338.1 Mb. All raw Illumina sequence data were passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts (L. Mingkun, A. Copeland, J. Han, unpublished). The following steps were then performed for assembly: (a) filtered Illumina reads were assembled using Velvet version 1.2.07 (6); (b) 1- to 3-kb simulated paired-end reads were created from Velvet contigs using wgsim version 0.3.0 (<https://github.com/lh3/wgsim>); and (c) Illumina reads were assembled with simulated read pairs using ALLPATHS-LG version r46652 (7). The final draft assembly contained 19 contigs in 18 scaffolds, totaling 4.2 Mb in size. The final assembly was based on 1,107.1 Mb of Illumina data, corresponding to 221.4× input read coverage.

Genes were identified using Prodigal (8), followed by a round of manual curation using GenePRIMP (9) for finished genomes and draft genomes in fewer than 20 scaffolds. The predicted coding sequences were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases.

The tRNAScanSE tool (10) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA (11). 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 version 1.1 (12). Additional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes platform developed by the JGI (13). Secondary metabolite gene clusters and possible encoded compounds were predicted with antiSMASH (14) and NaPDoS (15).

Using antiSMASH-3, the strain *Arthrobacter enclensis* NCIM 5488^T showed PKS 1 and PKS 3 secondary metabolite gene clusters encoding for polyketide synthases. Further, NaPDoS predicted the presence of gene clusters encoding for compounds such as nystatin and epothilone, along with fatty acid synthesis. These clusters are the first reported for any *Arthrobacter* sp. to date, and the results highlight the genome mining potential of the novel strain *Arthrobacter enclensis* NCIM 5488^T for natural products discovery research.

Nucleotide sequence accession numbers. This whole-genome shotgun project was deposited in DDBJ/ENA/GenBank under the accession number [LNQM000000000](https://www.ncbi.nlm.nih.gov/nuccore/LNQM000000000). The version described in this paper is the first version, LNQM01000000.

ACKNOWLEDGMENTS

This work was supported by the Council of Scientific and Industrial Research through project fund CSC 0407. A portion of this research was performed under the Genome Sequencing of Prokaryotic Type Strains at the DOE Joint Genome Institute under the Genomic Encyclopedia of Bacteria and Archaea (GEBA) project.

FUNDING INFORMATION

This work, including the efforts of Mahesh S. Dharne and Syed G. Dastager, was funded by the Council of Scientific and Industrial Research (CSIR) (CSC0407).

REFERENCES

1. Conn HJ, Dimmick I. 1947. Soil bacteria similar in morphology to *Mycobacterium* and *Corynebacterium*. *J Bacteriol* 54:291–303.

2. Mages IS, Frodl R, Bernard KA, Funke G. 2008. Identities of *Arthrobacter* spp. And *Arthrobacter*-like bacteria encountered in human clinical specimens. *J Clin Microbiol* 46:2980–2986. <http://dx.doi.org/10.1128/JCM.00658-08>.
3. Fenical W, Jensen PR. 2006. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat Chem Biol* 2:666–673. <http://dx.doi.org/10.1038/nchembio841>.
4. Dastager SG, Liu Q, Qin L, Tang SK, Krishnamurthi S, Lee JC, Li WJ. 2014. *Arthrobacter enclensis* sp. nov., isolated from sediment sample. *Arch Microbiol* 196:775–782. <http://dx.doi.org/10.1007/s00203-014-1016-9>.
5. Bennett S. 2004. Solexa Ltd. *Pharmacogenomics* 4:433–438. <http://dx.doi.org/10.1517/14622416.5.4.433>.
6. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
7. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 108:1513–1518. <http://dx.doi.org/10.1073/pnas.1017351108>.
8. 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>.
9. 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>.
10. 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>.
11. 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. <http://dx.doi.org/10.1093/nar/gkm864>.
12. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933–2935. <http://dx.doi.org/10.1093/bioinformatics/btt509>.
13. 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>.
14. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243. <http://dx.doi.org/10.1093/nar/gkv437>.
15. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. 2012. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7:e34064. <http://dx.doi.org/10.1371/journal.pone.0034064>.