

Draft Genome Sequence of the Novel Agarolytic Bacterium *Aquimarina agarilytica* ZC1

Bokun Lin,^{a,b} Guoyong Lu,^b Shengkang Li,^b Zhong Hu,^{a,b} and Honglin Chen^a

International Institute of Infection and Immunity, Shantou University Medical College, Shantou, Guangdong, China,^a and Department of Biology, College of Science, Shantou University, Shantou, Guangdong, China^b

The marine bacterium ZC1 is the type strain of the recently identified novel species *Aquimarina agarilytica*. It can produce multiple agarases. Here we report the draft genome sequence of strain ZC1 (4,253,672 bp, with a GC content of 32.8%) and major findings from its annotation. It is the first reported genome in the genus *Aquimarina*.

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance (3). Agar, which consists of agarose and agarpectin, is a component present in the cell walls of red algae. Agarase is a kind of enzyme that can degrade agarose and mainly comes from agarolytic bacteria. The agarase-producing bacterium ZC1 isolated from the surface of marine red alga is the type strain of the novel species *Aquimarina agarilytica* (7). The genome sequencing was performed to study the agar-degrading system of the novel type strain. Here, we present the draft genome sequence of strain ZC1. To our knowledge, this is the first genome report of the bacterial genus *Aquimarina*.

The genome of ZC1 was sequenced by a whole-genome shotgun strategy using the Illumina HiSeq 2000 at the Beijing Genomics Institute (BGI; Shenzhen, China). Genome sequences were assembled *in silico* using SOAPdenovo (6), resulting in 155 contigs with an N_{50} length of 56,586 bp. The protein-coding genes were predicted using Glimmer 3.02 (2). The functions of predicted protein-coding genes were then annotated through comparisons with the NCBI-NR, KEGG (4), and GO (1) databases. The annotation was done by using tRNAscan-SE 1.21 (8) to find tRNA and by using rRNAmmer 1.2 (5) to search rRNA.

The draft genome includes 4,253,672 bases with a GC content of 32.8% and contains 3,594 predicted genes and 3,456 predicted coding sequences (CDSs). An estimated 87.0% of nucleotides are predicted genes. There are 36 tRNAs and single-copy genes predicted for 5S, 16S and 23S rRNA. The 1,728 CDSs annotated by GO can be classified into 20 GO categories, and 2,047 CDSs can be annotated in the KEGG orthology system. A total of 246 CDSs were related to amino acid metabolism, while 204 CDSs were related to carbohydrate metabolism in the KEGG orthology system. A total of 207 CDSs were annotated as poorly characterized in the KEGG orthology system, indicating the novelty of the strain and its genes.

Forty CDSs were putative agarases in the KEGG orthology system. To our knowledge, this is the largest amount of predicted agarases in one bacterium. The lengths of these agarase genes ranged from 870 to 5,127 bp, and their protein sequence similar-

ities ranged from 29% to 70% (with 78% of them below 50%) to the submitted sequences in GenBank database. The low similarity indicated the novelty of the strain and its agarases. Gene cloning and expression of all 40 putative agarase genes are ongoing to study their roles in agarose degradation.

Nucleotide sequence accession number. The genome sequence has been deposited in GenBank under accession no. [AHHE00000000](https://www.ncbi.nlm.nih.gov/nuclseq/AHHE00000000).

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (no. 41076106), the Guangdong Natural Science Foundation (no. S2011030005257), the Science & Technology Project of Guangdong Province (no. 2009B030803051), and the Key Science and Technology Innovation Project for University by the Department of Education of Guangdong Province (no. CXZD1124).

REFERENCES

1. Ashburner M, et al. 2000. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29.
2. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673–679.
3. Henshaw J, et al. 2006. Family 6 carbohydrate binding modules in β -agarases display exquisite selectivity for the non-reducing termini of agarose chains. *J. Biol. Chem.* 281:17099–17107.
4. Kanehisa M, et al. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36:D480–D484.
5. Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of rRNA genes. *Nucleic Acids Res.* 35:3100–3108.
6. Li R, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
7. Lin B, et al. 2011. *Aquimarina agarilytica* sp. nov., a novel agarolytic species isolated from red alga. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/ijs.0.027136-0.
8. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of tRNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.

Received 24 February 2012 Accepted 28 February 2012

Address correspondence to Z. Hu, hzh@stu.edu.cn, or H. Chen, hlchen@hku.hk.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00311-12