Journals.ASM.org

Genome Sequence of *Thalassospira xiamenensis* Type Strain M-5

Qiliang Lai and Zongze Shao

Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography, State Oceanic Administration, Xiamen, People's Republic of China

Thalassospira xiamenensis **M-5^T was isolated from the surface water of a waste oil pool at the oil storage dock in the city of Xiamen, Fujian Province, China. Here, we present the draft genome of strain M-5^T, which contains 4,705,237 bp with a G+C content of 54.65% and contains 4,343 protein-coding genes and 46 tRNA genes.**

Thalassospira xiamenensis M-5^T (MCCC 1A00209^T = DSM 17429^T = CGMCC 1.3998^T) was isolated from the surface water of a waste oil pool at the oil storage dock in the city of Xiamen, Fujian Province, China, in December 2002 (1). As of now, 76 strains of this species have been isolated from marine environments in our lab (http://www.mccc.org.cn) (with >99% 16S rRNA gene sequence similarity). The genus *Thalassospira* was first proposed by Lopez-Lopez et al. (2) and contains five type strains (http://www.bacterio.cict.fr/t/thalassospira.html).

The genome sequence of *T. xiamenensis* M-5^T was determined by Shanghai Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China), using Solexa paired-end sequencing technology. A total of 3,325,692 paired-end reads (500-bp library) were generated to reach a 127-fold depth of coverage with an Illumina/Solexa IIx genome analyzer (Illumina, San Diego, CA), and the gaps among scaffolds were closed by custom primer walking or by PCR amplification followed by DNA sequencing. The genome of *T. xiamenensis* M-5^T consists of 76 contigs (>200 bp; N_{90} contigs = 30) of 4,705,237 bp and had an average G+C content of 54.65%. Automatic gene annotation was carried out by the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html), which was followed by manual editing. The genome contains 4,343 candidate protein-encoding genes (with an average size of 960 bp), giving a coding intensity of 88.7%. A total of 3,308 proteins could be assigned to cluster of orthologous groups (COG) families (3). Forty-six tRNA genes for 18 amino acids (Asn and Lys were absent) were identified. The proteins associated with amino acid transport and metabolism (COG initial, E) were the most abundant group of COG (422 open reading frames [ORFs], 12.7%), followed by those associated with transcription (K; 369 ORFs, 11.2%) and inorganic ion transport and metabolism (P; 285 ORFs, 8.6%).

Strain M-5^T cannot use *n*-alkane, naphthalene, phenanthrene, or pyrene as the sole carbon source for growth, though it was isolated from diesel fuel-degrading consortium. The genome sequence analysis revealed no alkane-degrading related monooxygenase present, but a gene encoding a ring hydroxylating dioxygenase awaits further characterization. The genome sequence of strain M-5^T and its curated annotation are important assets to better understand the physiology and metabolic potential of *Thalassospira xiamenensis* and will open up new opportunities in the function analysis of this species in oil-degrading consortia.

Nucleotide sequence accession number. The draft genome sequence of *Thalassospira xiamenensis* M-5^T has been deposited at GenBank under accession number AMRQ00000000 (chromosome).

ACKNOWLEDGMENTS

We acknowledge Qiang Li and his colleagues for genome analysis at Shanghai Majorbio Bio-pharm Technology Co., Ltd.

This work was financially supported by the Public Welfare Project of SOA (201005032) and the International Sci & Tech Cooperation Program of China (2010DFB23320).

REFERENCES

1. **Liu C, Wu Y, Li L, Ma Y, Shao Z.** 2007. Thalassospira xiamenensis sp. nov. and Thalassospira profundimaris sp. nov. Int. J. Syst. Evol. Microbiol. **57**:316–320.
2. **Lopez-Lopez A, et al.** 2002. Thalassospira lucentensis gen. nov., sp. nov., a new marine member of the alpha-Proteobacteria. Int. J. Syst. Evol. Microbiol. **52**:1277–1283.
3. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28**:33–36.

Received 28 September 2012 Accepted 1 October 2012

Address correspondence to Zongze Shao, shaozz@163.com.

doi:10.1128/JB.01904-12