JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of Marine-Derived *Streptomyces* sp. Strain AA0539, Isolated from the Yellow Sea, China

Zhi-Qiang Xiong and Yong Wang

Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Here, we report the draft genome sequence of *Streptomyces* sp. strain AA0539, isolated from marine sediment of the Yellow Sea, China. Its small genome (~5.8 Mb) contains large, unique genes and gene clusters for diverse secondary metabolites, suggesting great potential as a source for the discovery of novel natural products.

Streptomyces spp. are remarkably rich sources of natural products and still produce a larger number and variety of novel bioactive compounds than any other genus (13). Recent trends in drug discovery from natural sources emphasize investigation of marine *Streptomyces* spp., which have great potential to increase the number of natural products in clinical trials (12). *Streptomyces* sp. strain AA0539 was isolated from a marine sediment sample taken at a water depth of 50 m in the Yellow Sea (latitude, 35°12′125′′N; longitude, 119°41′881′′E). Its metabolites showed specific antitumor activity on a human ovarian carcinoma cell line, NIH:OVCAR-3. In the interest of the above-described information, we herein reported the draft genome sequence of *Streptomyces* sp. AA0539.

Raw data of the genome were generated by a Roche 454 Genome Sequencer FLX and assembled using Newbler 2.3. Putative protein-encoding sequences were identified by Glimmer 3.02 (2) and GeneMark (5). Functional annotation was based on BLASTP results with NR and KEGG databases. tRNA and rRNA genes were predicted with tRNAscan-SE (11) and RNAmmer (4), respectively. A total of 374,828 reads including up to 159.5 Mb were obtained, which represented a 27.2-fold coverage of the *Streptomyces* sp. AA0539 genome. The draft genome distributed in 57 contigs totaling 5,768,564 bp with an average GC content of 72.92% was significantly smaller than the genome size (8 to 12 Mb) of other reported *Streptomycetes* spp. (7), suggesting that the AA0539 genome is highly strain specific. This draft genome consisted of one linear chromosome with 3 rRNA operons, 55 tRNA genes, and 5,149 coding sequences (CDSs). For the CDSs, 3,814 proteins could be assigned to COG families (10), and 1,335 CDSs encode proteins with no match to any known proteins in the public databases. Additionally, the signal peptide cleavage sites, transmembrane topologies, and lipoproteins were predicted using SignalP 4.0 (8), TMHMM 2.0 (3), and LipoP 1.0 (9), respectively. We identified 433 proteins as secreted proteins, 1,184 proteins as transmembrane proteins, 1,055 lipoproteins, and 295 proteins as the transporter.

Twenty biosynthetic gene clusters for secondary metabolites (2 siderophores, 4 terpenes, 1 lantibiotic, 2 polyketide synthases [PKS], 3 nonribosomal peptide synthetases [NRPS], 2 nucleosides, 1 ectoine, and 5 hybrid NRPS/PKS) were identified by antiSMASH (6). For example, the predicted ectoine cluster is responsible for the biosynthesis of ectoine and 5-hydroxyectoine, which act as protectants against osmotic and heat stress in the cell (1). More than 168 diverse genes located in these secondary metabolic gene clusters were predicted using genome analysis, indicating high genomic synteny to those of various *Streptomyces* species. More importantly, many putative genes in these gene clusters showed low identity with the known ones, suggesting that AA0539 may be a potential producer of novel natural products. Hence, the unique genome information of *Streptomyces* sp. AA0539 will further promote the elucidation of chemical and genetic diversity of this strain for the discovery of novel gene clusters and natural products.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ALNP00000000. The version described in this paper is the first version, ALNP01000000.

ACKNOWLEDGMENTS

This work was supported by the National Basic Research Program of China (973 Program, grant no. 2012CB721104), the National High Technology Research and Development Program of China ("863" Program 2012AA02A701), the National Natural Science Foundation of China (grant no. 31170101 and 31100073), and the Major Projects of Knowledge Innovation Program of the Chinese Academy of Sciences (grant no. KSCX2-EW-J-12).

REFERENCES

1. **Bursy J, et al.** 2008. Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. Appl. Environ. Microbiol. **74**:7286–7296.
2. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
3. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**:567–580.
4. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
5. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26**:1107–1115.
6. **Medema MH, et al.** 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. **39**:W339–W346.
7. **Nett M, Ikeda H, Moore BS.** 2009. Genomic basis for natural product biosynthetic diversity in the actinomycetes. Nat. Prod. Rep. **26**:1362–1384.
8. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.
9. **Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC.** 2008. Methods for the

Received 2 September 2012 **Accepted** 21 September 2012

Address correspondence to Yong Wang, yongwang@sibs.ac.cn.

doi:10.1128/JB.01646-12

bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World J. Microbiol. Biotechnol. **24**:2377–2382.
10. **Reichard K, Kaufmann M.** 2003. EPPS: mining the COG database by an extended phylogenetic patterns search. Bioinformatics **19**:784–785.
11. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS Web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33**:W686–W689.
12. **Waters AL, Hill RT, Place AR, Hamann MT.** 2010. The expanding role of marine microbes in pharmaceutical development. Curr. Opin. Biotechnol. **21**:780–786.
13. **Xiong ZQ, Zhang ZP, Li JH, Wei SJ, Tu GQ.** 2012. Characterization of *Streptomyces padanus* JAU4234, a producer of actinomycin X2, fungichromin, and a new polyene macrolide antibiotic. Appl. Environ. Microbiol. **78**:589–592.