JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of the Marine *Streptomyces* sp. Strain AA1529, Isolated from the Yellow Sea

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 a *Streptomyces* strain, AA1529, isolated from marine sediment from the Yellow Sea. Its genome contains a subset of unique genes and gene clusters that encode diverse secondary metabolites, suggesting great potential as a source for the discovery of novel gene clusters and bioactive compounds.

Streptomyces bacteria are remarkably rich sources of natural products (NPs), accounting for the production of over two-thirds of the commercially available antibiotics in current use (1). This genus still produces a larger number and variety of novel bioactive compounds than any other genus (12). In the course of our screening program for new NPs from marine microorganisms, *Streptomyces* sp. strain AA1529 was isolated from a marine sediment sample taken at a water depth of 50 m (35°12′125″N, 119°41′881″E) in the Yellow Sea close to Rizhao City, Shangdong Province, China. Its metabolites had good specific antitumor activity on a human ovarian carcinoma cell line, NIH:OVCAR-3. Here we reported the draft genome sequence of *Streptomyces* sp. strain AA1529.

Raw genome data were generated by a Roche 454 Genome Sequencer FLX and assembled using the GS *de novo* Assembler software. A total of 391,115 reads including up to 170.7 Mb were obtained, which represented 23.4-fold coverage of the genome. The *Streptomyces* sp. strain AA1529 draft genome was distributed into 109 contigs totaling 7,285,977 bp with an average GC content of 72.8%. Putative protein-encoding sequences were identified using Glimmer (2) and GeneMark (5). Functional annotation was based on BLASTP results with the KEGG and NR databases. tRNA and rRNA genes were predicted with tRNAscan-SE (10) and RNAmmer (4), respectively. The signal peptide cleavage sites, transmembrane topologies, and lipoproteins were predicted by SignalP 4.0 (7), TMHMM 2.0 (3), and LipoP 1.0 (8), respectively. The draft genome consisted of one linear chromosome with 5 rRNA operons, 58 tRNA genes, and 6,334 coding sequences (CDSs). For the CDSs, 4,235 proteins could be assigned to Clusters of Orthologous Groups families (9) and 576 CDSs encode proteins with no match to any known proteins in the current public databases, suggesting that the AA1529 genome is highly strain specific. As for the subcellular localization of the proteins, we identified 445 proteins as secreted proteins, 1,402 proteins as transmembrane proteins, and 301 proteins as transporters.

Twenty-three secondary-metabolite (2 siderophores, 3 terpenes, 2 lantibiotics, 6 polyketide synthases [PKS], 4 nonribosomal peptide synthetases [NRPS], and 6 hybrid NRPS-PKS) biosynthetic gene clusters were identified by antiSMASH (6). One hundred sixteen diverse secondary metabolic genes located in various gene clusters on the AA1529 genome were predicted by genome analysis, suggesting high genomic synteny to those of various *Streptomyces* species. Many putative genes involved in antibiotic biosynthesis showed low identity with the known ones, suggesting that AA1529 may be a potential producer of novel NPs. Additionally, some antimicrobial peptides predicted by the CAMP database (11) had significant DNA and amino acid sequence variations from their orthologs from other *Streptomyces* species. Thus, mining of the *Streptomyces* sp. strain AA1529 genome will further elucidate the chemical and genetic diversity of this strain for the discovery of novel gene clusters and bioactive compounds.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. ALAP00000000. The version described in this paper is the first version, ALAP01000000.

ACKNOWLEDGMENTS

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

REFERENCES

1. **Bentley SD, et al.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417**:141–147.
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**:W339–346.
7. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.
8. **Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC.** 2008. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World J. Microbiol. Biotechnol. **24**:2377–2382.

Received 11 July 2012 **Accepted** 27 July 2012

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

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

doi:10.1128/JB.01247-12

9. **Reichard K, Kaufmann M.** 2003. EPPS: mining the COG database by an extended phylogenetic patterns search. Bioinformatics **19**:784–785.
10. **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.
11. **Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S.** 2010. CAMP: a useful resource for research on antimicrobial peptides. Nucleic Acids Res. **38**:D774–D780.
12. **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.