Draft Genome Sequence of Marine *Streptomyces* sp. Strain W007, Which Produces Angucyclinone Antibiotics with a Benz[*a*]anthracene Skeleton

Song Qin,^a Hongyu Zhang,^{a,c} Fuchao Li,^b Benwei Zhu,^a and Huajun Zheng^d

Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, China^a; Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China^b; Graduate University of Chinese Academy of Sciences, Beijing, China^c; and Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China^d

A series of angucyclinone antibiotics have been isolated from marine *Streptomyces* sp. strain W007 and identified. Here, a draft genome sequence of *Streptomyces* sp. W007 is presented. The genome contains an intact biosynthetic gene cluster for angucyclinone antibiotics, which provides insight into the combinatorial biosynthesis of angucyclinone antibiotics produced by marine streptomycetes.

Streptomyces sp. strain W007, isolated from marine sediment from Jiaozhou Bay, produces a series of angucyclinone antibiotics; 3-hydroxy-1-keto-3-methyl-8-methoxy-1,2,3,4-tetrahydro-benz[*a*]anthracene showed high cytotoxicity against human lung adenocarcinoma cell line A549 (15). The compound contains a methoxyl group at C-8 and no keto groups at C-7 and C-12, different from oviedomycin (8), jadomycin B (5), simocyclinone (4, 13), and angucyclines Sch 47554 and Sch 47555 (1). Herein, we report the draft genome sequence of marine *Streptomyces* sp. W007.

The nucleotide sequence was determined using a 454 GS FLX sequencer, and assembly was performed using the GS De Novo Assembler software version 2.3 (11). A total of 536,432 reads including up to 186,808,079 bp were obtained, which represented a 20.6-fold coverage of the genome. Finally, we obtained the *Streptomyces* sp. W007 draft genome of 9,057,348 bp distributed in 284 contigs with a G + C content of 71.36%.

Putative protein-encoding sequences were predicted using Glimmer (3) and GeneMark (10). Functional annotation was based on BLASTP results with the Kyoto Encyclopedia of Genes and Genomes (KEGG) and NR databases. tRNA genes were directly predicted with tRNAscan-SE (9). The signal peptide cleavage sites, transmembrane topologies, and lipoproteins were predicted using SignalP 3.0 with hidden Markov models (2), PSORTb version 3.0.2 (14), and LipoP 1.0 (6), respectively.

The genome consists of one linear chromosome with 4 rRNA operons, 64 tRNA genes, and 7,904 protein-encoding genes (coding sequences [CDSs]). Among the CDS products, 4,898 proteins could be assigned to clusters of orthologous groups (COG) families (12). Three hundred eight hypothetical proteins have no match to any known proteins in the databases. At least 764 multigene (paralog) families corresponding to 2,968 predicted proteins were identified. As for the subcellular localization of the proteins, 461 proteins were identified as secreted proteins, 1,266 proteins as transmembrane proteins, and 149 proteins as lipoproteins. Similar to the first reported marine *Streptomyces* genome (7), the genome of *Streptomyces* sp. W007 carries genes encoding the ABC transporter ATP-binding proteins and two-component regulatory systems widely distributed on the chromosome, a feature which could be related to marine environmental adaptation.

Genome analysis revealed a number of genes related to biosyn-

thesis of secondary metabolites. The gene cluster for angucyclinone antibiotics contains genes encoding polyketide synthesis cyclase, 3-oxoacyl-(acyl carrier protein) synthase II, beta-ketoacyl synthase, cyclase, short-chain dehydrogenase, carboxyl transferase, and the acyl coenzyme A (acyl-CoA) carboxylase alpha subunit. This cluster shows high-level homologies to biosynthetic gene clusters for oviedomycin (8), jadomycin B (5), simocyclinone (4, 13), and angucyclines Sch 47554 and Sch 47555 (1). However, there are special genes for O-methyltransferase, dehydrase, and reductase in the gene cluster for 3-hydroxy-1-keto-3-methyl-8-methoxy-1,2,3,4-tetrahydro-benz[*a*]anthracene in *Streptomyces* sp. W007. There are several genes related to glycosyl biosynthesis in the genome, including genes for dTDP-glucose 4,6-dehydratase, dTDP-6-deoxy-L-hexose 3-O-methyltransferase, glucose-1-phosphate thymidyltransferase, and dTDP-4-keto-6-deoxy-L-hexose-2,3-dehydratase, which show high-level homologies to glycosyl biosynthesis genes for simocyclinone (4, 13) and angucyclines Sch 47554 and Sch 47555 (1). In this case, marine *Streptomyces* sp. W007 has the potential to produce angucyclinones with glycosyl, which provides us information to isolate novel compounds on purpose.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AGSW000000000](https://www.ncbi.nlm.nih.gov/nuccore/AGSW000000000). The version described in this paper is the first version, AGSW01000000.

ACKNOWLEDGMENTS

This work was supported by grants from the CAS International Innovation Partnership Program: Typical Environmental Process and Effects on Resources in Coastal Zone Area, the National Natural Science Foundation of China (grant no. 31000037), the Outstanding Young Scholar fellowship

Received 11 December 2011 Accepted 4 January 2012

Address correspondence to Song Qin, sqin@yic.ac.cn, or Fuchao Li, lifuchao@qdio.ac.cn.

S.Q. and H.Z. contributed equally to this work.

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

doi:10.1128/JB.06701-11

program of Shandong province (grant no. JQ200914), the Natural Science Foundation of Shandong province (grant no. ZR2009EQ004), the Knowledge Innovation Program of the Chinese Academy of Sciences (grant no. KSCX2-EW-G-12), and the Foundation of the Key Laboratory of Marine Bioactive Substance and Modern Analytical Techniques, SOA (grant no. MBSMAT-2010-07).

REFERENCES

1. Basnet DB, et al. 2006. Angucyclines Sch 47554 and Sch 47555 from *Streptomyces* sp. SCC-2136: cloning, sequencing, and characterization. *Mol. Cells* 22:154–162.
2. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340:783–795.
3. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27: 4636–4641.
4. Galm U, et al. 2002. Cloning and analysis of the simocyclinone biosynthetic gene cluster of *Streptomyces antibioticus* Tü 6040. *Arch. Microbiol.* 178:102–114.
5. Han L, Yang K, Ramalingam E, Mosher RH, Vining LC. 1994. Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. *Microbiology* 140:3379–3389.
6. Juncker AS, et al. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12:1652–1662.
7. Li F, et al. 2011. Draft genome sequence of the marine bacterium *Streptomyces griseoaurantiacus* M045, which produces novel manumycin-type antibiotics with a pABA core component. *J. Bacteriol.* 193:3417–3418.
8. Lombó F, Braña AF, Salas JA, Méndez C. 2004. Genetic organization of the biosynthetic gene cluster for the antitumor angucycline oviedomycin in *Streptomyces antibioticus* ATCC 11891. *ChemBiochem* 5:1181–1187.
9. 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.
10. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* 26:1107–1115.
11. Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
12. 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.
13. Trefzer A, et al. 2002. Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents Chemother.* 46:1174–1182.
14. Yu NY, et al. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615.
15. Zhang H, et al. 2011. A new anthracene derivative from marine *Streptomyces* sp. W007 exhibiting highly and selectively cytotoxic activities. *Mar. Drugs* 9:1502–1509.