

Draft Genome Sequence of the Marine *Streptomyces* sp. Strain PP-C42, Isolated from the Baltic Sea[∇]

Longjiang Fan,¹ Yun Liu,¹ Zefeng Li,¹ Heike I. Baumann,³ Katrin Kleinschmidt,³ Wanzhi Ye,² Johannes F. Imhoff,³ Michael Kleine,⁴ and Daguang Cai^{2*}

James D. Watson Institute of Genome Sciences & Institute of Crop Science, Zhejiang University, Hangzhou 310029, China¹; Department of Molecular Phytopathology, Christian-Albrechts-Universität zu Kiel, D-24118 Kiel, Germany²; Department of Marine Microbiology, Leibniz-Institute of Marine Science (IFM-GEOMAR), Düsternbrooker Weg 20, D-24105 Kiel, Germany³; and Planton GmbH, Am Kiel-Kanal 44, D-24106 Kiel, Germany⁴

Received 15 April 2011/Accepted 4 May 2011

***Streptomyces*, a branch of aerobic Gram-positive bacteria, represents the largest genus of actinobacteria. The streptomycetes are characterized by a complex secondary metabolism and produce over two-thirds of the clinically used natural antibiotics today. Here we report the draft genome sequence of a *Streptomyces* strain, PP-C42, isolated from the marine environment. A subset of unique genes and gene clusters for diverse secondary metabolites as well as antimicrobial peptides could be identified from the genome, showing great promise as a source for novel bioactive compounds.**

Streptomyces, a branch of aerobic Gram-positive bacteria, represents the largest genus of actinobacteria, with more than 900 described species. Streptomycetes are characterized by the formation of mycelia and spores during their life cycle and by a complex secondary metabolism. They produce more than two-thirds of the clinically used natural antibiotics (4, 11) and represent an important source of novel bioactive compounds. Furthermore, many enzymes produced by streptomycetes are important for food manufacturing (3) and for diverse industrial applications (7). *Streptomyces* strains have been isolated and characterized from a large variety of habitats (10, 13). Although more than five *Streptomyces* genomes have been completely sequenced in recent years, numerous genome sequencing projects with different *Streptomyces* species are still ongoing. Streptomycetes have linear chromosomes (approximately 8 to 12 Mb) with a high G+C content (11). More than 20 diverse secondary metabolic gene clusters in their genome have been described to date (11, 12).

Streptomyces sp. strain PP-C42 was isolated from the surface layer of a sediment core taken at a water depth of 241 m from the Gotland Deep in the Baltic Sea. The sampling was taken through a small gravity corer during an expedition with the research vessel *Alkor* AL156 in the year 2000.

Raw data of the genome were generated from two independent sequencing lanes using Illumina GA II and assembled with the Velvet program (18). The released genome sequence of the *Streptomyces* strain IFO 13350 (12) served as a reference. GeneMarkS (2), tRNAscan-SE (14), and RNAmmer (8) were utilized to predict protein-coding genes, tRNAs, and rRNAs, respectively. The GSP software (<http://gsizepred.sourceforge.net>) was used to estimate the genome size of the strain (5, 9).

The draft genome sequence of strain PP-C42 comprises 7,167,114 bases representing approximately 74.5% of the 9.6-Mb estimated size of the PP-C42 genome. The genome of strain PP-C42 has a high G+C content of 72.5%. The draft genome sequence contains 4,410 open reading frames (ORFs), 62 tRNAs, and 24 rRNAs. Of 4,410 ORFs, 2,774 genes have orthologs in *Streptomyces* strain IFO 13350 (12) (BLASTP < 1e−5), while 1,076 ORFs were not found in any of the five released genome sequences of other *Streptomyces* strains (1, 11) and 1,068 ORFs did not give any hits in the current public databases. This may be a reflection of a high degree of the strain specificity of the PP-C42 genome. So far, 19 diverse secondary metabolic genes have been identified; these genes are located on the PP-C42 genome in various gene clusters, which exhibit high genomic synteny to those of various *Streptomyces* species. Also, a set of hits was retrieved (BLASTP < 1e−5) from various antimicrobial peptide databases (6, 15, 16, 17) but with striking sequence variations at both DNA and amino acid levels when compared to their orthologs from other *Streptomyces* species. Thus, the unique genome information provided by the draft sequence of PP-C42 has great importance for basic as well as applied microbial genomic researches.

Nucleotide sequence accession numbers. This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AEWS00000000. The version described in this paper is the first version under accession number AEWS01000000.

This project was supported by the Bundesministerium für Bildung und Forschung (BMBF), Germany (grant 0315231A, B), and the Ministerium für Wissenschaft, Wirtschaft und Verkehr des Landes Schleswig-Holstein (grant 122-08-002). We thank DAAD (grant D/08/01773, 4) and the China Scholarship Council (grant A/10/00701) for providing the scholarship reward as well as international exchange grants.

We thank Jun Wang for his help with the Solexa sequencing and Katharina Peetz for her technical support.

* Corresponding author. Mailing address: Department of Molecular Phytopathology, Christian-Albrechts-University of Kiel, Hermann, Rodewald Str. 9, D-24118 Kiel, Germany. Phone: 49-431-8803215. Fax: 49-431-8801583. E-mail: dcai@phytomed.uni-kiel.de.

[∇] Published ahead of print on 13 May 2011.

REFERENCES

1. **Anderson, A. S., and E. M. Wellington.** 2001. The taxonomy of *Streptomyces* and related genera. *Int. J. Syst. Evol. Microbiol.* **51**:797–814.
2. **Besemer, J., A. Lomsadze, and M. Borodovsky.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**:2607–2618.
3. **Blättel, V., et al.** 2009. A lytic enzyme cocktail from *Streptomyces* sp. B578 for the control of lactic and acetic acid bacteria in wine. *Appl. Microbiol. Biotechnol.* **83**:839–848.
4. **Ceylan, O., G. Okmen, and A. Ugur.** 2008. Isolation of soil *Streptomyces* as source antibiotics active against antibiotic-resistant bacteria. *Eurasia J. Bio-Sci.* **2**:73–82.
5. **Chor, B., D. Horn, N. Goldman, Y. Levy, and T. Massingham.** 2009. Genomic DNA k-mer spectra: models and modalities. *Genome Biol.* **10**:R108.
6. **Gueguen, Y., et al.** 2006. PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. *Dev. Comp. Immunol.* **30**:283–288.
7. **Ladjama, A., Z. Taibi, and A. Meddour.** 2007. Production of pectinolytic enzymes using *Streptomyces* strains isolated from palm grove soil in Biskra area (Algeria). *Afr. Crop Sci. Conf. Proc.* **8**:1155–1158.
8. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**:3100–3108.
9. **Marçais, G., and C. Kingsford.** 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770.
10. **Moran, M. A., L. T. Rutherford, and R. E. Hodson.** 1995. Evidence for indigenous *Streptomyces* populations in a marine environment determined with a 16S rRNA probe. *Appl. Environ. Microbiol.* **61**:3695–3700.
11. **Nett, M., H. Ikeda, and B. S. Moore.** 2009. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**:1362–1384.
12. **Ohnishi, Y., et al.** 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190**:4050–4060.
13. **Pathom-Aree, W., et al.** 2006. Diversity of actinomycetes isolated from Challenger Deep sediment (10,898 m) from the Mariana Trench. *Extremophiles* **10**:181–189.
14. **Schattner, P., A. N. Brooks, and T. M. Lowe.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**:W686–689.
15. **Thomas, S., S. Karnik, R. S. Barai, V. K. Jayaraman, and S. Idicula-Thomas.** 2010. CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* **38**:D774–780.
16. **Wang, C. K., Q. Kaas, L. Chiche, and D. J. Craik.** 2008. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* **36**:D206–210.
17. **Wang, G., X. Li, and Z. Wang.** 2009. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **37**:D933–937.
18. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.