JB Accepts, published online ahead of print on 13 May 2011
J. Bacteriol. doi:10.1128/JB.05097-11
Copyright © 2011, American Society for Microbiology and/or the Listed Authors/Institutions. All Rights Reserved.

JB Accepts published online ahead of print

The draft genome sequence of a marine *Streptomyces* sp. strain PP-C42 isolated from the Baltic Sea

Formatted: Numbering: Continuous

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]*

1. James D. Watson Institute of Genome Sciences & Institute of Crop Science, Zhejiang University, Hangzhou 310029, China
2. Molecular Phytopathology, Christian-Albrechts-Universität zu Kiel, D-24118 Kiel, Germany
3. Marine Microbiology, Leibniz-Institute of Marine Science (IFM-GEOMAR), Düsternbrooker Weg 20, D-24105 Kiel, Germany
4. Planton GmbH, Am Kiel-Kanal 44, D-24106 Kiel, Germany

Correspondence to	Prof. Dr. Daguang Cai
Phone:	+49-431-8803215
Fax:	+49-431-8801583
E-mail:	dcai@phytomed.uni-kiel.de
Address:	Department of Molecular Phytopathology, Christian-Albrechts-University of Kiel, Hermann, Rodewald Str. 9, D-24118 Kiel, Germany

Key words: Genome sequencing, *Streptomyces*, antimicrobial peptides, marine isolate

JB Accepts published online ahead of print

Abstract

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 (AMPs) could be identified from the genome, showing great promise as a source for novel bioactive compounds.

JB Accepts published online ahead of print

Results and discussion

Streptomyces, a branch of aerobic Gram-positive bacteria represents the largest genus of actinobacteria with over 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 over two-thirds of the clinically used natural antibiotics (4, 11) and represent an important source for 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 the last years, numerous genome sequencing projects with different *Streptomyces* species are still ongoing. Streptomycetes have linear chromosomes (approx. 8–12 Mb) with a high G+C content (11). More than 20 diverse secondary metabolic gene clusters in their genome have been described, so far (11, 12). *Streptomyces* strain PP-C42 was isolated from the surface layer of a sediment core taken in 241 m water depth of the Gotland Deep in the Baltic Sea. The sampling was taken through a small gravity corer during the expedition with the research vessel Alkor AL156 in the year 2000.

Raw data of the genome were generated from two independent lanes using Illumina GA II and assembled with 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 genome size of the strain (5, 9).

The draft genome sequence of strain PP-C42 comprises 7167114 bases representing approx. 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

JB Accepts published online ahead of print

draft genome sequence contains 4410 open reading frames (ORFs), 62 tRNAs and 24 rRNAs. Of 4410 ORFs, 2774 genes have orthologs in *Streptomyces* strain IFO 13350 (12) (BLASTP <1e-5) while 1076 ORFs were not found in any of the five released genome sequences of other *Streptomyces* strains (1, 11) and 1068 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, nineteen diverse secondary metabolic genes have been identified, which locate 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 for 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 the accession number: AEWS01000000.

Acknowledgements

This project was supported by the Bundesministerium für Bildung und Forschung (BMBF), Germany (grant number 0315231A, B) and the Minsterium für Wissenschaft, Wirtschaft und Verkehr des Landes Schleswig-Holstein (grant number 122-08-002). Authors thanks DAAD (grant number D/08/01773, 4) and China Scholarship Council (grant number A/10/00701) for providing the scholarship reward as well as international exchange grants. Authors thank Jun Wang for his help in Solexa sequencing and Ms. Katharina Peetz for her

JB Accepts published online ahead of print

technical support.

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. **Blattel, V., K. Wirth, H. Claus, B. Schlott, P. Pfeiffer, and H. Konig.** 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 BioSci **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., J. Garnier, L. Robert, M. P. Lefranc, I. Mougenot, J. de Lorgeril, M. Janech, P. S. Gross, G. W. Warr, B. Cuthbertson, M. A. Barracco, P. Bulet, A. Aumelas, Y. Yang, D. Bo, J. Xiang, A. Tassanakajon, D. Piquemal, and E. Bachere.** 2006. PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. Dev Comp Immunol **30:**283-288.
7. **Ladjama, A., Taibi.Z, and Meddour.A.** 2007. Production of pectinolytic enzymes using Streptomyces strains isolated from palm grove soil in Biskra area(Algeria). African Crop Science Conference Proceedings **8:**1155-1158.
8. **Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100-3108.
9. **Marcais, 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., J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A. Yamashita, M. Hattori, and S. Horinouchi.** 2008. Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350. J Bacteriol **190:**4050-4060.
13. **Pathom-Aree, W., J. E. Stach, A. C. Ward, K. Horikoshi, A. T. Bull, and M. Goodfellow.** 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.

JB Accepts published online ahead of print