

Draft Genome Sequence of *Streptomyces globisporus* C-1027, Which Produces an Antitumor Antibiotic Consisting of a Nine-Membered Eneidyne with a Chromoprotein

Lifei Wang, Songmei Wang, Qing He, Tengfei Yu, Qinglian Li, and Bin Hong

Key Laboratory of Biotechnology of Antibiotics of Ministry of Health, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

***Streptomyces globisporus* C-1027 is the producer of antitumor antibiotic C-1027, a nine-membered eneidyne-containing compound. Here we present a draft genome sequence of *S. globisporus* C-1027 containing the intact biosynthetic gene cluster for this antibiotic. The genome also carries numerous sets of genes for the biosynthesis of diverse secondary metabolites.**

Streptomyces globisporus C-1027, isolated from a soil sample collected in the Qian-jiang area of China (2), produces an extremely potent antitumor antibiotic, C-1027. As a member of eneidyne family, C-1027 is currently undergoing phase II clinical trial in China (8). Here, we present a draft genome sequence of *S. globisporus* C-1027.

The nucleotide sequencing was performed by BGI (Shenzhen, China) using an illumina/GA sequencer. Short reads were assembled by SOAP *de novo* (3), and protein coding sequences (CDSs) were predicted by Glimmer, version 3.0 (1). Gene functional annotation was based on BLASTP results determined with the KEGG, COG, Swiss-Prot, NT, and NR databases. tRNA genes were predicted with tRNAscan-SE (7), and rRNA sequences were found by alignment with an rRNA pool.

A total of 1,242.47 Mb of data were generated, which represented 169.37-fold coverage of the genome. The draft genome sequence of *S. globisporus* C-1027 contains 7,693,617 bp with a GC content of 71.63% distributed over 84 scaffolds containing 278 contigs. We identified 7,231 putative protein CDSs, accounting for 88.22% of the genome, as well as 56 tRNA genes and five rRNA operons in our draft genome. The intact biosynthetic gene cluster for C-1027, which had been cloned and sequenced previously (4), was located on scaffold 34, with 99% identity to the sequence of the reported C-1027 biosynthetic genes. In addition, scaffold 34 disclosed a further 6,264 bp upstream of the C-1027 gene cluster in which a pair of genes composing a primase/helicase-like gene and a putative replication initiation gene were predicted. Together with three other pairs of these genes previously annotated adjacent to the C-1027 gene cluster, there are a total of four pairs of genes involved in the linear plasmid replication. The coding proteins of these genes are highly homologous to the plasmid-type DNA primase/replication proteins of plasmid pSLA2-L from *S. rochei* (5) and plasmid SCP1 from *S. coelicolor* (6), suggesting that the biosynthetic gene cluster of C-1027 is possibly located on a giant linear plasmid at least 92 kb in length. Genome analysis also revealed a number of genes related to biosynthesis of diverse secondary metabolites besides C-1027, including putative nonribosomal peptide synthetase (NRPS) genes, polyketide synthase (PKS) genes, NRPS-PKS hybrid genes, terpene cyclase genes, and genes for lantibiotic biosynthesis. The genome sequence of *S. globisporus* C-1027 will aid our further analysis and understanding of C-1027 biosynthesis-regulatory mechanisms and discovery of new natural products by uncovering cryptic metabolic pathways.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. [AJUO00000000](https://www.ncbi.nlm.nih.gov/nuccore/AJUO00000000). The version described in this paper is the first version, deposited under [AJUO01000000](https://www.ncbi.nlm.nih.gov/nuccore/AJUO01000000).

ACKNOWLEDGMENTS

We thank BGI (Shenzhen, China) for Solexa shotgun sequencing, assembly, and annotation of the genome.

This work was supported by the National Natural Science Foundation of China (31170042 and 30973668), Ministry of Science and Technology of China (2012ZX09301002-001-016, 2009ZX09501-008, and 2010ZX09401-403), and Beijing Natural Science Foundation (5102032).

REFERENCES

- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673–679.
- Hu JL, et al. 1988. A new macromolecular antitumor antibiotic, C-1027. I. Discovery, taxonomy of producing organism, fermentation and biological activity. *J. Antibiot.* (Tokyo) 41:1575–1579.
- Li R, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
- Liu W, Shen B. 2000. Genes for production of the eneidyne antitumor antibiotic C-1027 in *Streptomyces globisporus* are clustered with the *cagA* gene that encodes the C-1027 apoprotein. *Antimicrob. Agents Chemother.* 44:382–392.
- Mochizuki S, et al. 2003. The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol. Microbiol.* 48:1501–1510.
- Redenbach M, Bibb M, Gust B, Seitz B, Spychaj A. 1999. The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. *Plasmid* 42:174–185.
- 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.
- Shao RG, Zhen YS. 2008. Eneidyne anticancer antibiotic lidamycin: chemistry, biology and pharmacology. *Anticancer Agents Med. Chem.* 8:123–131.

Received 8 May 2012 Accepted 21 May 2012

Address correspondence to Bin Hong, binhong69@hotmail.com.

L.W. and S.W. contributed equally to this article.

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

doi:10.1128/JB.00797-12