Data in Brief

Complete genome sequence of a natural compounds producer, *Streptomyces violaceus* S21

Jiafang Fu\(^1,1\), Chuanqing Zhong\(^b,1\), Zhilong Zhao\(^c\), Gongli Zong\(^a\), Guangxiang Cao\(^a,***\)

\(^a\) Shandong Medicinal Biotechnology Centre, Shandong Academy of Medical Sciences, Jinan, China

\(^b\) Shandong Jianzhu University, Jinan, China

\(^c\) College of Drug, Linyi University, Linyi, China

**ARTICLE INFO**

**Keywords:**
Streptomyces violaceus
Complete genome sequence
Natural compounds
Anthracycline antibiotics

**ABSTRACT**

The complete genome sequence of *Streptomyces violaceus* strain S21, a valuable natural compounds producer isolated from the forest soil, is firstly presented here. The genome comprised 7.91M bp, with a G + C content of 72.65%. A range of genes involved in pathways of secondary product biosynthesis were predicted. The genome sequence is available at DDBJ/EMBL/Genbank under the accession number CP020570. This genome is annotated with 6856 predicted genes identifying the natural product biosynthetic gene clusters in *S. violaceus*.

**Specifications**

| Organism/cell line/tissue | Streptomyces violaceus |
|---------------------------|------------------------|
| Sequencer or array type   | Illumina Hiseq4000, PacBio RS II |
| Data format               | Raw and processed |
| Experimental factors      | Microbial strains |
| Experimental features     | Complete genome sequence of *Streptomyces violaceus* strain S21, assembly and annotation |
| Consent                   | N/A |
| Sample source location    | The Seabed sludge (Shandong Province, China) (36°67′5.7″ N, 122°98′28.0″ E). |

**1. Direct link to deposited data**

The complete genome sequences can be found at the site [https://www.ncbi.nlm.nih.gov/nuccore/CP020570](https://www.ncbi.nlm.nih.gov/nuccore/CP020570). The raw data can be found at the site [http://dx.doi.org/10.17632/66hfym3csx.1](http://dx.doi.org/10.17632/66hfym3csx.1).

**2. Introduction**

Natural products from actinomycetes have been the major sources for clinical antibiotics, along with numerous other useful compounds including antineoplastic, antiparasitic, insecticidal and phytocidal drugs. *Streptomyces violaceus* is one of the great potential producers of natural compounds. *S. violaceus* was first isolated and classified in the 1960s [1]. Following its discovery, different kinds of anthracycline antibiotics for cancer treatment were isolated from *S. violaceus*, but side effects like cardiotoxicity have limited their clinical use [2–5]. Development of new anthracyclines with less cardiotoxicity and improved therapeutic efficacy is required. In addition, amylase inhibitors, extracellular polysaccharide and thrombolytic actinoprotease were also isolated from *S. violaceus* [6–8]. *S. violaceus* was also used to develop new useful natural products recently [9–10].

**3. Experimental design, materials and methods**

Strain S21 was isolated from the Seabed sludge in Shandong, China. Strain S21 is a valuable producer of many natural compounds, including anthracycline antibiotics, amylase inhibitors and extracellular polysaccharide. Analysis of the genome of strain S21 was carried out in order to reveal the biosynthetic gene clusters of natural compounds. *S. violaceus* S21 was cultured in Tryptic Soy Broth (OXOID, UK) medium to obtain mycelium, then Genomic DNA was extracted using Genomic DNA Purification Kit (Promega, USA). Both the PE300 DNA library and 10-kb DNA library were constructed, after the quality of...
DNA sample was analyzed using a NanoDrop 2000 Spectrophotometer (Thermo Scientific, USA). DNA sequencing was performed using an Illumina Hiseq4000 platform and a PacBio RS II platform at Beijing Genomics Institute (Shenzhen, China). The cleaned reads were de novo assembled with SPAdes [11], then polished with SSPACEStandard and GapFiller to get scaffolds [12–13]. The genome was annotated using the Prokaryotic Genome Annotation Pipeline (PGAP) version 3.2 software on NCBI. Additional gene prediction was performed by the RASTtk server [14]. SEED viewer was used for assignment of the predicted genes to functional categories [15].

4. Data description

After quality control, about 1.50 Gb of data was obtained from the Illumina Hiseq platform, and about 0.81 Gb of data was obtained from the PacBio RS II platform. A total of 7,916,045 bp genome sequence with an average GC content of 72.65% was assembled. The genome was predicted to contain 6856 genes, including 6571 coding sequences, 65 tRNAs, 18 rRNAs (5S, 16S, and 23S), 3 ncRNAs, and 199 pseudo genes. Most of the annotated genes determined amino acids and derivative synthesis (667), carbohydrate metabolism (409), cofactor, vitamin, prosthetic group and pigment formation (357), protein metabolism (349), and fatty acid, lipid and isoprenoid (149) (Fig. 1).

About 79 gene clusters involved in the pathways for the secondary products biosynthesis were predicted in the genome of Strain S21 using antiSMASH [16]. Further studies of the genes involved in the biosynthesis of anthracycline antibiotics, amylase inhibitors and extracellular polysaccharide are necessary.

5. Nucleotide sequence accession numbers

The nucleotide sequence of the S. violaceus S21 genome has been deposited in Genbank under the accession number CP020570.

Conflict of interest

The authors declare that there is no conflict of interests on the work published in this paper.

Acknowledgements

This work was supported, in part, by the Natural Science Foundation of Shandong (ZR2016CP22) and the Innovation Project of Shandong Academy of Medical Sciences (201604).

References

[1] E. Munoz, M.V. Elorza, J.R. Villanueva, Classification of a microorganism with lytic action as a type of Streptomyces violaceus, Microbiol. Exp. 17 (1964) 195–208.
[2] O. Joho, T. Ishikura, A. Yoshimoto, T. Takeuchi, Anthracycline metabolites from Streptomyces violaceus A262. I. Isolation of antibiotic-blocked mutants from Streptomyces violaceus A262. J. Antibiot. 44 (10) (1991) 1110–1117.
[3] O. Joho, T. Yoshioka, T. Takeuchi, A. Yoshimoto, Isolation of new anthracyclines 10-O-rhodosaminyl beta-rhamdoesonicine and beta-isorhodosonicine from mild-acid treated culture of obelinycin-producing Streptomyces violaceus, J. Antibiot. 50 (6) (1997) 522–525.
[4] Y. Miyamoto, O. Joho, Y. Naganamatsu, A. Yoshimoto, Cloning and characterization of a glycosyltransferase gene involved in the biosynthesis of anthracycline antibiotic beta-rhodosomycin from Streptomyces violaceus, FEMS Microbiol. Lett. 206 (2) (2002) 163–168.
[5] Y. Miyamoto, S. Ohta, O. Joho, Y. Naganamatsu, A. Yoshimoto, Production of a new hybrid anthracycline 4-O-methylepseudomycin by heterologous expression of dmkR in epelmycin-producing Streptomyces violaceus, J. Antibiot. 53 (8) (2000) 828–836.
[6] P. Manivasan, P. Visvanar, J. Venkatkum, S. K. Silvukums, S. K. Kim, Production and characterization of an extracellular polysaccharide from Streptomyces violaceus MM72, Int. J. Biocatal. 59 (2013) 29–38.
[7] V. Mohanaravina, C.D. Sabathra, S. Yogesh, A. Govindaraj, S.N. Jenimah, In vitro thrombolytic potential of antiprotopeptide from marine Streptomyces violaceus VITYGM, Cardiovasc. Hematol. Agents Med. Chem. 14 (2016) 120–124.
[8] N. Shovava, Amylase inhibitors from Streptomyces lucensis VKPM Ac-1743 and Streptomyces violaceus VKPM Ac-1734, Prikl. Biokhim. Mikrobiol. 51 (1) (2015) 46–52.
[9] J.R. Doroghazi, J.C. Albright, A.W. Goring, K.S. Ju, R.R. Haines, K.A. Tchalekian, D.F. Labeled, N.L. Kelleher, W.W. Metcalf, A roadmap for natural product discovery based on large-scale genomics and metabolomics, Nat. Chem. Biol. 10 (11) (2014) 963–968.
[10] K.S. Ju, J. Gao, J.R. Doroghazi, K.K. Wang, C.J. Thibodeaux, K.K. Metzger, J. Fudala, J. Su, J.R. Zhang, J. Lee, J.P. Cioni, B.S. Evans, R. Hirota, D.P. Labeda, W.A. van der Donk, W.W. Metcalf, Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes, Proc. Natl. Acad. Sci. U. S. A. 112 (39) (2015) 12175–12180.
[11] A. Bankevich, S. Nukf, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lein, S.I. Nikolenko, S. Pham, A.D. Pyrbelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (5) (2012) 455–477.
[12] M. Boerger, C.V. Henkel, H.S. Jensen, D. Butler, W. Pirvano, Scaffolding pre-assembled contigs using SPAGE, Bioinformatics 27 (4) (2011) 578–579.
[13] M. Boerger, W. Pirvano, Toward almost closed genomes with GapFiller, Genom. Biol. 13 (6) (2012) R56.
[14] T. Breetin, J.J. Davis, T. R.A. Edwards, S. Gerdes, G.J. Olsen, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, J.A. Thomas and, R. Stevens, V. Vonstein, A.R. Wattam, F. Xia, RAST: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes, Sci. Rep. 5 (2015) 8365.
[15] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. R.A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A.R. Wattam, F. Xia, R. Stevens, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST), Nucleic Acids Res. 42 (Database issue) (2014) D206–D214.
[16] T. Weber, K. Blin, S. Duddela, D. Krug, H.U. Kim, R. Brucoli, S.Y. Lee, M.A. Fischbach, R. Muller, W. Wohleben, R. Breitling, E. Takano, M.H. Medema, antimSMASH 3.0: a comprehensive resource for the genome mining of biosynthetic gene clusters, Nucleic Acids Res. 43 (W1) (2015) W237–W243.