Synthetic and Systems Biotechnology 1 (2016) 56–65

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: keaipublishing.com/synbio

Genome mining reveals the biosynthetic potential of the marine-derived strain *Streptomyces marokkonensis* M10

Liangyu Chen [a], Ying-Mi Lai [b], Yu-Liang Yang [b], Xinqing Zhao [c,*]

[a] School of Life Science and Biotechnology, Dalian University of Technology, Dalian 116024, China
[b] Agricultural Biotechnology Research Center, Academia Sinica, Taipei 11529, Taiwan
[c] State Key Laboratory of Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form 27 February 2016
Accepted 27 February 2016
Available online

Keywords:
Streptomyces marokkonensis
Genome mining
Secondary metabolites
Polyketide synthase
Polyene macrolides

ABSTRACT

Marine streptomycetes are rich sources of natural products with novel structures and interesting biological activities, and genome mining of marine streptomycetes facilitates rapid discovery of their useful products. In this study, a marine-derived *Streptomyces* sp. M10 was revealed to share a 99.02% 16S rDNA sequence identity with that of *Streptomyces marokkonensis* Ap1^T, and was thus named *S. marokkonensis* M10. To further evaluate its biosynthetic potential, the 7,207,169 bps of *S. marokkonensis* M10 genome was sequenced. Genomic sequence analysis for potential secondary metabolite-associated gene clusters led to the identification of at least three polyketide synthases (PKSs), six non-ribosomal peptide synthases (NRPSs), one hybrid NRPS-PKS, two lantibiotic and five terpene biosynthetic gene clusters. One type I PKS gene cluster was revealed to share high nucleotide similarity with the candicidin/FR008 gene cluster, indicating the capacity of this microorganism to produce polyene macrolides. This assumption was further verified by isolation of two polyene family compounds PF1 and PF2, which have the characteristic UV adsorption at 269, 278, 290 nm (PF1) and 363, 386 and 408 nm (PF2), respectively. *S. marokkonensis* M10 is therefore a new source of polyene metabolites. Further studies on *S. marokkonensis* M10 will provide more insights into natural product biosynthesis potential of related streptomycetes. This is also the first report to describe the genome sequence of *S. marokkonensis*-related strain.

1. Introduction

Streptomyces are Gram-positive bacteria that are prolific sources of secondary metabolites and contribute to the vast majority of the microbial-derived natural products.[1] Extensive studies have been performed on marine-derived streptomycetes due to the diverse chemical structures and important biological activities of their secondary metabolites, which serve as sources for novel antibiotics to combat with the emerging antibiotic-resistant pathogens.[2]

Genome sequencing studies have demonstrated greater biosynthetic potential of streptomycetes than previously expected from a genetic perspective. Such studies were initially carried out on *Streptomyces coelicolor* A3(2) and *S. avermitilis*,[3,4] where many biosynthetic gene clusters associated with the secondary metabolites were unveiled, indicating that even such relatively well-explored streptomycetes species have the potential to yield much more new compounds than have been discovered.[5] In recent years, our knowledge on natural product biosynthesis potential of streptomycetes has been enriched by the complete genome sequencing of more and more *Streptomyces* species,[6–8] and a lot of genome sequencing projects of various *Streptomyces* species that are still ongoing. Genome mining, one of the bioinformatics-based approaches for natural product discovery, has been developed based on these genome sequencing projects and has been applied to discover chemical structures of novel unidentified molecules.[5,9–11]

Exploration of polyketides and some peptide antibiotics especially benefits from genome information and genome mining approach due to the presence of polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) and lantibiotic synthases, which sequentially assemble small carboxylic acid and amino acid into products like an assembly line.[12] The corresponding biosynthetic genes are usually clustered together with regulatory and resistance elements, transport systems and some other relevant

* Corresponding author. State Key Laboratory of Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China. Tel.: +86 21 34206673; fax: +86 21 34208028.
E-mail address: xqzhao@sjtu.edu.cn (X. Zhao).
Peer review under responsibility of KeAi Communications Co., Ltd.

http://dx.doi.org/10.1016/j.synbio.2016.02.005

functional genes. Consequently, the biosynthetic products could be predicted easily with bioinformatics approach from the genome sequence and gene functions.[13]

In our studies searching for novel antibiotics from marine-derive streptomycetes, *S. xinghaiensis* and *S. xiaopingdaonensis* (previously named *S. sulphureus* L180)[14,15] have been characterized. Genome sequencing of these two strains revealed various possible biosynthetic gene clusters of secondary metabolites.[16,17] Here we report the draft genome sequence of the marine-derived streptomycete M10, which was selected due to its strong antifungal activity. The secondary metabolic biosynthetic gene clusters of M10 genome were analyzed via genome mining, which guided the discovery of two polyene compounds.

2. Materials and methods

2.1. Strains and culturing conditions

The M10 strain was isolated from the marine sediment collected in Dalian, China, and cultured on Bennett's agar for 2 weeks at 28 °C. The strain was preserved in our lab as a glycerol stock at –70 °C and at the China General Microbiological Culture Collection (CGMCC, accession number 7143). TSB medium (BD Difico, USA) was used for seed culture and A1 agar (soluble starch 10 g/L, yeast extract 4 g/L, peptone 2 g/L, artificial sea salt 28 g/L) was used for bioactive secondary metabolites extraction and analysis.

Candida albicans (CGMCC 2.538) and *Fulvia fulva* (kindly provided by Prof. Qiu Liu from Dalian Nationalities University, China) were employed as indicator pathogens which were maintained on Yeast Extract Peptone Dextrose (YPD, yeast extract 5 g/L, peptone 10 g/L, glucose 20 g/L) and Potato Dextrose Agar (PDA, potato 200 g/L, glucose 2 g/L, $(NH_4)_2SO_4$ 1.0 g/L, $MgSO_4$ 1.0 g/L, agar 1.75 g/L) slants at 4 °C, respectively.

2.2. Analysis of the 16S rDNA sequence

M10 was cultured on TSB agar at 30 °C for two weeks for a morphological observation and in TSB broth at 30 °C for 4 days to harvest mycelia for genomic DNA extraction and PCR amplification of 16S rDNA gene sequence was performed according to the method described previously.[17] The sequencing result was aligned via the NCBI BLAST program (http://blast.ncbi.nlm.nih.gov/) and the EzTaxon-e database (http://eztaxon-e.ezbiocloud.net/)[18] to choose the closely related strains to identify the 16S rDNA gene sequence similarities among them.

2.3. Genome sequencing, annotation and analysis

The draft genome sequence of M10 was obtained by a combination of Roche/454 pyrosequencing and Illumina/Solexa sequencing to afford an assembly with scaffolds, which was performed by Beijing Genome Institute (BGI) in Shenzhen, China. The paired-end reads generated by Illumina sequencing were assembled by SOAPdenovo1.05.[19] Coding sequences were predicted by Prodigal.[20] Functional assignment of coding genes was obtained by performing a sequence similarity search with BLAST against the Clusters of Orthologous Groups (COG, http://www.ncbi.nlm.nih.gov/COG/) reference database, and functional gene annotation was based on BLASTP with the KEGG databases. The Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AMZL00000000 and the version described in this paper is the first version, AMZL01000000.

The natural product-associated gene clusters were further identified and categorized by Antibiotics & Secondary Metabolite Analysis SHell (antiSMASH 3.0.2) and Artemis Release 12.0 software by BLASTP alignment searching for key words such as PKS, NRPS, ironophore, lantibiotic, terpene, etc. against the model natural product domains and genes in the NCBI database.[21,22] The upstream and downstream regions of core genes were subsequently investigated and putative biosynthetic gene clusters were proposed. The alignment between two genomes or gene clusters were achieved by Double ACT v2 and visualized by the software SACT_v9 to assist the reassembling of the scaffolds.

Table 1
Primers used for RT-PCR analysis.

Primers	Sequence (5′—3′)
marRI-F	CGCAGAGTTCGGAGGACGAG
marRI-R	ACCCCCGTAATGCGAAACGAAG
marRII-F	CATGACCCTGCTCCCCGAAC
marRII-R	CAGTTCCTTCAACCGGTGCGC
marRIII-F	GACTGGCCACCACCATCGAG
marRIII-R	GAAACGGTCCAGCACGTCGTG
marRIV-F	GAGCTGACCGCTCACTCCTTC
marRIV-R	GGTTGGTGTTCCAGCACGCC

2.4. Total RNA extraction and gene expression analysis by RT-PCR

M10 was inoculated on A1 agar plates and cultured for 2, 4 or 6 days for RNA extraction. Total RNA was extracted by using RNAsimple Total RNA Extraction Kit (Tiangen Biotech Inc., China) and RNA reverse transcription for cDNA was performed using the PrimeScript RT Reagent Kit (Dalian Takara Inc., China). The transcription of four regulatory genes *marRI*, *marRII*, *marRIII* and *marRIV* was selected to evaluate with PCR primers listed in Table 1. PCR reaction conditions were as follows: 4 min at 94 °C for one cycle, followed by 1 min at 90 °C, 30 s at 58 °C and 2 min at 72 °C for 40 cycles, and finally one cycle for 10 min at 72 °C.

2.5. Purification of polyene molecules from M10

M10 was inoculated on A1 agar plates and totally 630 plates were cultured at 28 °C for one week. Both of the mycelia and agar were cut into small pieces (about 3 × 3 cm) and extracted three times firstly with ethyl acetate (EtOAc) and then with n-butyl alcohol (BuOH) overnight to afford the EtOAc extract (2.1 g) and BuOH extract (5.6 g). The EtOAc and BuOH extracts were then fractionated on Sephadex LH-20 (Sigma) column separately eluting with EtOAc and MeOH at a flow rate of 1.5 mL/4 min. Each fraction was dried and resuspended in MeOH (1–2 mL), 100 μL of which were added to the lawns of *C. albicans* and *F. fulva* for the antifungal bioassay. Bioactive fractions were then inspected by MALDI-TOF for molecular signatures. The targeted fractions were subsequently fractionated by flash C18 column chromatography eluted with 0%, 20%, 50%, 80% and 100% MeOH (MeOH/water, v:v). The samples were finally analyzed and purified using water/acetonitrile gradient and monitoring at 254, 280, 300, 360 and 380 nm on HPLC system with Discovery HS C18 columns (250 × 4.6 mm, 5 μm, 1 mL/min; 250 × 10 mm, 10 μm, 5 mL/min).

2.6. MALDI-TOF analysis of bioactive fractions

MALDI-TOF analysis of Sephadex LH-20 fractions was performed in positive ion mode with a mass range of 200–1500 Da on Bruker Autoflex Speed MALDI-TOF mass spectrometer (Bruker Daltonics Inc., USA). In general, 1 μL saturated matrix solution of universal MALDI matrix (1:1 mixture of 2-5-dihydroxybenzoic acid and α-cyano-4-hydroxy-cinnamic acid, Sigma-Aldrich, USA) in 78%/21%/1% (v/v) acetonitrile/water/formic acid and 1 μL sample (dissolved in methanol) were mixed together and spotted on the MALDI MSP 96 anchor plate until dried. Then the plate was subjected