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

Liangyu Chen, Ying-Mi Lai, Yu-Liang Yang, Xinqing Zhao

**A B S T R A C T**

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, 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 candidicin/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.

**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

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. 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.

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*, 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. 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 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.

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. The corresponding biosynthetic genes are usually clustered together with regulatory and resistance elements, transport systems and some other relevant functional elements.
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 −80°C and at the China General Microbiological Culture Collection (CGMCC, accession number 7143). TSB medium (BD Difco, 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, (NH₄)₂SO₄ 1.0 g/L, MgSO₄ 1.0 g/L, agar 17.5 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 2 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 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.

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% (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

| Table 1 | Primers used for RT-PCR analysis. |
|---------|----------------------------------|
| Primers | Sequence (5′–3′)                  |
| marRI-F | CGCAGAGTCGGAGGACACGAG            |
| marRI-R | ACCCCCGTAGCGGAAACGAG             |
| marRII-F | CATGACCTGTCCCTCCGAAC            |
| marRII-R | CGTCTCTTCACCGTGTCGG             |
| marRIII-F | GACCTGCAACCACCCAGTCGAG         |
| marRIII-R | GAAACGTCGCCAGCTCGTG              |
| marRIV-F | GAGCTGACCCCTCCTCTCCT             |
| marRIV-R | GGTGGTGTTCCTCAGCCAGGC           |
to the MALDI-TOF mass spectrometer for MS acquisition and data were analyzed to describe the chemical profiles of the fractions by FlexAnalysis 2.0 software.

3. Results

3.1. General features of the genome of strain M10

The partial 16S rDNA gene sequence of the M10 strain, 1422 bps in length, was deposited in the GenBank nucleotide database with an accession number of JX397867. Phylogenetic analyses indicated that M10 belongs to the genus *Streptomyces* and shared the highest gene identity of 16S rDNA (99.02%) with the type strain *S. marokkonensis* Ap1. It was reported that *S. marokkonensis* Ap1 was isolated from rhizosphere soil of the indigenous Moroccan plant *Argania spinosa* L. and a producer of antifungal polyene macrolides. Therefore this is the first report that *S. marokkonensis*-related strain was isolated from marine environment. Thus M10 was named *S. marokkonensis* M10.

The draft genome sequence of *S. marokkonensis* M10 is 7,207,169 bps in length, with an average G + C content of 73.46% (Fig. 1A). The assembly consists of 552 contigs (500 bps) with an N50 size of 31 kb. The genome of *S. marokkonensis* M10 consists of one putative linear chromosome with 6482 coding sequences (CDSs), representing approximately 88.62% of the whole genome, and the average gene length is 985 bps (Table 2). In addition, 4456 of the total CDSs give no direct BLAST functional results, but showed some similarities to known genome sequences. Additionally, 96 CDSs with high identities are just assigned as hypothetical protein. Such a large percentage of uncertain functional CDSs indicate the high potential of this strain in producing more novel proteins and compounds.

According to genome-wide alignment, *S. albus* J1074 has the highest similarity to *S. marokkonensis* M10. As 76 scaffolds still remain after the automatic assembling, genome sequence of *S. albus* J1074, which shows high similarity to that of *S. marokkonensis* M10, is employed to help manually re-assemble the genome (Fig. 1B). Although *S. marokkonensis* M10 does not show the highest similarity to *S. albus* J1074 through 16S rDNA comparison, the anti-SMASH annotation reveals that a large part of gene clusters from M10 show high similarities to those from the genome of *S. albus* J1074. However, despite the fact that over 6.4 Mb nucleotide sequences from the *S. marokkonensis* M10 genome could find their corresponding positions in the genome of *S. albus* J1074, there are still over 0.8 Mb sequences which are quite different from that of *S. albus* J1074. These extra sequences imply that *S. marokkonensis* M10 may have specific biosynthetic potentials in its genome. With the information obtained from the alignment, the fragmented *pks1* gene cluster, which was scattered in over nine scaffolds, as well as the nrps5 gene cluster, which is separated into two scaffolds, is re-assembled to enable further analysis. The gene clusters of *nrps6, lan1* and *pks-nrps2* in *S. marokkonensis* M10 are observed to be absent in the genome of *S. albus* J1074, implying that *S. marokkonensis* M10 has more biosynthetic capacity than *S. albus* J1074, which will be further discussed below.

3.2. Biosynthetic gene cluster associated with secondary metabolites

A combination of manual and automated methods was initially applied to annotate the *S. marokkonensis* M10 genome to predict the biosynthetic gene clusters. The full length of these gene clusters is estimated as 1029.3 kb, dedicating 14.28% of its genome (Table 2). Each putative ORF (Open Reading Frame) was compared with a typical representative library of the known PKS and NRPS domains as well as other known modular-type biosynthetic genes. The biosynthetic potential of *S. marokkonensis* M10 was compared with that of the model type strains *S. coelicolor* A3(2), and several other commonly studied strains, including *S. griseus* IFO 13350, *S. avermitilis*, *S. albus* J1074, *S. tropica* CNB-440, *N. farcinica* IFM 10152 and *S. marokkonensis* M10 consists of over nine scaffolds, as well as the fragmented *nrps6, lan1* and *pks-nrps2* in *S. marokkonensis* M10 are observed to be absent in the genome of *S. albus* J1074, implying that *S. marokkonensis* M10 has

### Table 2

Predicted gene clusters in *S. marokkonensis* M10 genome data and comparison with those of other actinomycetes strains.

| Organism                  | Size (Mb) | SMC No. | PKS Type I | PKS Type II | PKS Type III | NRPS | Hybrid PKS/NRPS | Non-NRPS siderophore |
|---------------------------|-----------|---------|------------|-------------|--------------|------|----------------|----------------------|
| *S. marokkonensis* M10    | 7.33      | ≥26     | ≥3         | –           | ≥6           | 1    | 2              |                      |
| *S. coelicolor* A3(2)     | 8.72      | ≈8      | 2          | 2           | 3            | 3 – 1| 1              |                      |
| *S. griseus* IFO 13350    | 8.55      | 34      | 7          | 2           | 2            | 6    | 1              |                      |
| *S. avermitilis*          | 9.03      | 38      | 2          | –           | 1            | 3    | 3              | 2                    |
| *S. albus* J1074          | 6.84      | 22      | 1          | 2           | 1            | 3    | 4              | 1                    |
| *S. tropica* CNB-440      | 5.18      | 17      | 4          | 1           | 1            | 1    | 7              |                      |
| *N. farcinica* IFM 10152  | 6.01      | – 12    | 1          | 1           | 1            | 1    | 7              |                      |

Mb, megabases; SMC, secondary metabolite gene clusters; ND, not determined; –, not applicable.

The putative natural product biosynthetic gene clusters were summarized in Table 3. The *pks1* cluster which belonged to the type I PKS is the largest biosynthetic gene cluster in the genome of *S. marokkonensis* M10. The *pks1* gene cluster shares high nucleotide sequence similarity with the candidin FR008 biosynthetic gene cluster from *Streptomyces* sp. FR-008. Another *pks1* cluster contains multiple genes similar to that in the Herboxidiene biosynthetic gene cluster. Other natural product related gene clusters were also predicted based on the genome sequence, including six NRPS, three non-NRPS, three lantibiotics and five terpenoid-related biosynthetic gene clusters. The six NRPS gene clusters were mined via NRPS database for candidate NRPS-derived peptides, including one mannopeptimycin-like, one antimycin-like and one desotamides-like cluster, respectively. We also identified a frotalalides-like cluster containing hybrid PKS-NRPS, and one terpenoid cluster similar to gremacene D.

The deduced functions of proteins encoded by *pks1* biosynthetic gene cluster were listed in Table 4, and the *pks1* gene cluster clearly distinguished itself from the candidin gene cluster from *S. griseus* IMRJ 3570 but somehow is quite similar to the FR-008 gene cluster from *Streptomyces* sp. FR-008. Due to the high degree of sequence conservation of the PKS domains, the alignments with the genome of *S. albus* J1074 and the macrolide biosynthetic gene cluster of *S. albus* J1074 were applied to re-assemble the scaffolds. Based on the alignment results, the missing MarB encoding sequence is obtained by assembling five scaffolds and found to be highly similar to FscB. Similarly, other two genes encoding type I polyketide synthase MarC and MarE were also assembled. By analyzing the alignment results, some differences of MarC and MarE from their
Fig. 1. Linear map of the assembled chromosome of *S. marokkonensis* M10 (illustrated as a circular one) by alignment and the none-assembled scaffolds (illustrated as a linear one) generated from DNAPlotter software. (A) The outer ring and the top line show the locations of natural product-associated gene clusters. The middle circle and middle line show the scale in bps, with 0 representing the origin of replication. The center ring and the third line represent a normalized plot of GC content. The inner ring and the bottom line show a normalized plot of GC skew. (B) The alignment between the assembled chromosome of *S. marokkonensis* M10 (top line) and that of *S. albus* J1074 (bottom line) generated from SACT_v9 software.
corresponding proteins FscC and FscE encoded by FR-008 gene cluster were observed, which will be further discussed below.

Partial sequence of MarC was identified from the fragmented small scaffolds, but the final total coverage is only 88%. The module prediction showed that at least 6 domains should be missing if the modules 5–10 are truly inside MarC like in FscB. Actually, the existence of the module 7 is still doubtful based on the available information. FscC is reported to be responsible for the assembling

Table 3
Characteristic of gene clusters in S. marokkonensis M10.

| No. | Cluster designation | Predicted product | Type | Size (kb) | Gene location | References |
|-----|---------------------|-------------------|------|----------|---------------|------------|
| 1   | pks1                | FR-008-like polyene macrolides | Type I PKS | 147.9 | Gene 2271–2287 | Scaffold 53, 44, 58, 41, 60, 55, 37 |
| 2   | pks2                | Herboxidiene like | Type III PKS | 41.1 | Gene 4825–4868 | Gene 4799–4824 |
| 3   | pks3\*              | Unknown           | Type I PKS | 29.5 | Gene 4113–4137 |            |
| 4   | pks4\#              | Unknown           | Type I PKS | 22.8 | Gene 4239–4247 |            |
| 5   | pks5\#              | Unknown           | Type I PKS | 19.2 | Gene 4256–4268 |            |
| 6   | nprs1               | Mannopeptimycin-like hexapeptide | NRPS | 62.2 | Gene 2638–2677 |            |
| 7   | nprs2               | Unknown           | NRPS | 61.2 | Gene 3871–3916 |            |
| 8   | nprs3               | Antimycin like    | NRPS | 43.4 | Gene 2249–2264 |            |
| 9   | nprs4               | Desotamide like   | NRPS | 58.5 | Gene 2337–2356 | Gene 3198–3218 |
| 10  | nprs5               | Unknown peptide   | NRPS | 20.0 | Gene 3173–3196, 5234, 4269–4285, 2085–2086 |

| 11  | nprs6\#             | Unknown peptide   | NRPS | 68.2 | Gene 3259–3292 |            |
| 12  | nprs7\#             | Unknown peptide   | NRPS | 43.4 | Gene 4100–4112 |            |
| 13  | pks-nprsi           | Frontalalides like | PKS-NRPS | 49.4 | Gene 3716–3749 |            |
| 14  | pks-nprsi2          | Unknown           | PKS-NRPS | 47.2 | Gene 3557–3579 |            |
| 15  | lam1                | Lantipeptide      | Lantibiotic | 25.3 | Gene 3367–3392 |            |
| 16  | lam2                | Thiopptide-lantipeptide | Lantibiotic | 32.5 | Gene 5459–5488 |            |
| 17  | Lam3\*              | Lantipeptide-NRPS | Lantipeptide-NRPS | 45.7 | Gene 2220–2248 |            |
| 18  | terp1               | Hogne-like terpene | Terpene | 26.6 | Gene 3654–3679 |            |
| 19  | terp2               | Unknown terpene   | Terpene | 20.3 | Gene 3914–3928 |            |
| 20  | terp3               | Germacrene D      | Terpene | 20.2 | Gene 4291–4312 |            |
| 21  | terp4               | Unknown terpene   | Terpene | 21.2 | Gene 4604–4621 |            |
| 22  | terp-bac            | Isorenieratene like | Terpena-bacteriocin | 31.1 | Gene 4961–4955 |            |
| 23  | sid1                | Desferrioxamine_B | NRPS-independent siderophore | 12.0 | Gene 0673–0684 |            |
| 24  | sid2                | Unknown siderophore | NRPS-independent siderophore | 15.1 | Gene 3403–3416 |            |
| 25  | lin                 | Unknown           | Other | 17.9 | Gene 4319–4335 |            |
| 26  | las                 | Unknown           | Other | 20.9 | Gene 2547–2563 |            |
| 27  | bac1                | Bacteriocin       | Other | 5.5  | Gene 3587–3590 |            |
| 28  | bac2                | Bacteriocin       | Other | 10.9 | Gene 3068–3079 |            |
| 29  | ect                 | Ectoine           | Other | 10.1 | Gene 6040–6049 |            |

* One end of the gene cluster is not completed.
* Both ends of the gene cluster are not completed.

Table 4
Deduced functions of proteins encoded by pks1 biosynthetic gene cluster.

| Protein | Amino acids | Deduced function | The most similar sequence | Identities/similarity | Accession number |
|---------|-------------|------------------|--------------------------|-----------------------|------------------|
| PabAB   | 723         | ADC synthase     | PabAB                    | 98%/98%               | AAQ2560          |
| PabC    | 257         | ADC lyase        | PabC                     | 96%/97%               | AAQ2560          |
| MarO    | 400         | FAD-dependent monoxygenase | FscO | 99%/98% | AAQ2549          |
| MarA    | 1744        | Type I polyketide synthase | FscA | 96%/96% | AAQ2561          |
| MarB    | 1806        | Type I polyketide synthase | FscB | 96%/96% | AAQ2565          |
| MarC    | 11006       | Type I polyketide synthase | FscC | 97%/97% | AAQ2568          |
| MarD    | 9472        | Type I polyketide synthase | FscD | 97%/97% | AAQ2568          |
| MarE    | 7848        | Type I polyketide synthase | FscE | 97%/97% | AAQ2568          |
| MarF    | 2040        | Type I polyketide synthase | FscF | 94%/94% | AAQ2566          |
| MarRI   | 148         | LuxR family transcriptional regulator | FscRI | 98%/97% | KP_007749177 |
| MarRII  | 942         | LuxR family transcriptional regulator | FscRII | 97%/98% | AAQ2553          |
| MarRIII | 1017       | LuxR family transcriptional regulator | FscRIII | 97%/98% | AAQ2553          |
| MarRIV  | 963         | LuxR family transcriptional regulator | FscRIV | 97%/98% | AAQ2554          |
| MarMI   | 458         | Glycocolysotransferase | FscMI | 99%/99% | KP_007749173 |
| MarMII  | 352         | GDP-ketosugar aminotransferase | FscMII | 99%/99% | AAQ2556          |
| MarMIII | 349        | GDP-mannose-4, 6-dehydratase | FscMIII | 99%/99% | AAQ2569          |
| MarP    | 393         | Cytochrome P450 monoxygenase | PimG | 99%/99% | CAC20928         |
| MarFE   | 64          | Ferredoxin       | FscFE                    | 100%/100%            | KP_007749170 |
| MarTE   | 256         | Type II thioesterase | FscTE | 98%/98% | AAQ2559          |
| MarTI   | 335         | ABC transporter  | FscTI                    | 97%/98% | KP_007749166    |
| MarTII  | 262         | ABC transporter  | FscTII                   | 90%/95%              | AAQ2563          |

* Query cover: 88%.
** Query cover: 78%.
of six (out of totally seven) conjugated double bonds. From the genomic information of M10, there might be five modules in MarC, indicating that the polyene molecules produced by \textit{S. marokkonensis} M10 might be smaller than candicidin (Fig. 2).

After manual assembly, there is still a 5 kb gap in MarE, about one fifth of the length of its closely related homologous protein FscE. It was found that the actual module arrangement of MarE is different to some extent from that of FscE, which can lead to changes in the structure of the chemical product. The order of module 17 and module 18 is different from that in FscE, and while module 17 possesses an ER domain, module 18 does not. Module 19 has one more KR domain, which is also different from FscE. It is therefore deduced that the possible product has a hydroxyl group instead of the ketone group in position of C-5 in the structure of FR-008. In addition, the KR domain in module 15 is not observed, and it is still not clear how this will affect the structure of the product.

3.3. Expression analysis of four regulatory genes

To check whether the putative polyene biosynthetic gene cluster is indeed transcriptionally active, four transcriptional regulatory genes \textit{marRI}, \textit{marRII}, \textit{marRIII} and \textit{marRIV} were investigated using the RNA samples collected 2, 4 and 6 days post-inoculation. As illustrated in Fig. 3, \textit{marRII} and \textit{marRIII} were expressed in the 2nd day but \textit{marRI
and marRI didn’t show any expression in the three selected time points, implying that marRI and marRII may play important roles in the pks1-directed biosynthesis of polyene molecules.

3.4. Genome mining guided natural product discovery of polyene macrolides

The excellent antifungal activity of S. marokkonensis M10 against C. albicans and F. fulva together with the pks1 cluster indicated the potential capacity of S. marokkonensis M10 to produce antifungal polyene compounds, which promoted us to further study the related metabolites. After primary purification, the fractions with the antifungal activity were further subjected to purification, and the fractions eluted from Sephadex LH-20 column were then collected and inspected for compounds with characteristic UV chromophores associated with polyene compounds. Two polyene family fractions (PF), with the typical polyene UV absorption at 269, 278, 290 nm (PF1) and 363, 386 and 408 nm (PF2), respectively, were then detected in fractions Fr.09, Fr.10 and Fr.11 from the BuOH crude extract fractionated by Sephadex LH-20 column (Fig. 4A,B). The PFs were collected based on time interval every minute and then tested the antifungal activity, and the results were shown in Fig. 4C. The capacity of S. marokkonensis M10 to produce ployene compounds were further exemplified by the isolation of compound 9-04 from PF1. In the 1H-NMR spectrum of 9-04, the peaks at 85.3–6.5 ppm indicated that there are several conjugated double bonds and the number was further deduced to be 3 or 4 by compared with other polyene compounds (Fig. 5A), while the number of conjugated double bonds for PF2 was predicted to be 7. Meanwhile, the 13C-NMR spectrum of 9-04 at δ176 ppm revealed a lactone moiety was involved in the chemical structure of 9-04 and verified the assumption that compound 9-04 could be a triene macrolide (Fig. 5B). On the other side, due to the low production level of PF2 and its instability to the light, PF2 could only be deduced to be heptaene based on the UV spectrum and raw 1H-NMR spectrum. The UV spectrum also gave us a hint that the MarC may carry 6 modules to synthesize PF2 with total seven double bonds. Regarding the only gene cluster able to synthesize the heptaene in the genome, the link between PF2 and the pks1 gene cluster would be highly possible.

4. Discussion

The increasing fungal infections and multi-drug resistance of fungal pathogen intensively promoted us to seek for novel, safe and effective antifungal antibiotics, thus led to the isolation of a marine-derived streptomycte S. marokkonensis M10, which is a similar strain of S. marokkonensis Ap1 (99.02% 16S rDNA sequence similarity). The type strain S. marokkonensis Ap1 was reported to be an antifungal polyene macrolide (pentaenes) producer. However, the genome information and secondary metabolic biosynthetic gene clusters of this species have not been reported previously, which promoted us to perform a genome sequencing of S. marokkonensis M10 and predict the bioactive secondary metabolites via the related gene clusters before traditional natural product isolation.

Multiple putative metabolic biosynthesis gene clusters, which are associated with the production of PKS, NRPS, lantibiotic and terpenoid, were predicted from the genome of S. marokkonensis M10, indicating the excellent natural product biosynthetic potential of S. marokkonensis M10, and guided us to specifically search for the molecules related with gene clusters. The four LuxR family of transcriptional regulatory genes (fscRI, fscRII, fscRIII, and fscRIV), were reported to maintain the stability of the extremely long mRNAs of the large PKS genes in FR-008 biosynthesis process and the disruption of fscRI, fscRII and partial fscRII led to the absence of FR-008.46 The regulatory genes (marRI, marRII, marRIII and marRIV) were deduced to act as the same roles in pks1 gene cluster. The expression of marRI and marRII by RT-PCR showed that the gene cluster should be transcriptionally active and strongly supported the assumption that S. marokkonensis M10 would express the polyene molecules. In addition, the active prior transcription of marRI and marRII in the 2nd day rather than in the 4th and 6th days was in accordance with the exhibition of antifungal activity of S. marokkonensis M10 from the 2nd day.

The analysis of type I PKS gene cluster pks1 successfully directed the isolation of two polyene molecule families PF1 and PF2 due to the characteristic UV profiles of polyene. Polyene natural
products, with unrivaled track record as antifungal antibiotics, were initially recognized as amphotericin B and nystatin,\textsuperscript{38,39} and subsequently candidin, as well as a series of newly isolated polyene compounds such as bahamaolides, marinisporolides, wortmannilactones and takanawaenes\textsuperscript{40–43} were also reported. The UV absorption spectrum of a polyene antibiotic usually enables it to be classified not only as a polyene, but also more specifically as a triene, tetraene, pentaene, hexaene or heptaene based on the UV shift caused by the increasing number of conjugated double bonds. The molecules involved in PF1 and PF2 should be classified as triene and heptaene, respectively. The partial structure elucidation of compound 9-04 in PF1 by \textsuperscript{1}H-NMR also verified the analysis above and more specifically, 9-04 was identified as triene macrolides in consideration of the lactone deduced by \textsuperscript{13}C-NMR spectrum.

Although the precise organization of \textit{pks1} gene cluster was not defined to date due to the low production amount and the instability to the light, the unique polyene-related gene cluster of \textit{pks1} in \textit{S. marokkonensis} M10 genome was very likely correlated to the production of polyene molecules within PF2. However, further genetic studies are necessary to prove that the putative polyene gene

---

\textbf{Fig. 4.} Identification of the polyene families in the BuOH layer crude extract of \textit{S. marokkonensis} M10 on A1 agar plate. (A) HPLC profile of polyene families PF1 (red-boxed) and PF2 (yellow-boxed). (B) Characteristic UV adsorption spectrum of PF1 and PF2. (C) The antifungal activity against \textit{C. albicans} of the M10 culture broth (left) and the HPLC fractions (middle and right). The number of the fraction corresponds to the time (min) in the HPLC profile in Fig. 4A.
Fig. 5. (A) $^1$H-NMR and (B) $^{13}$C-NMR spectrum of compound 9-04 in PF1. Compound was dissolved in C5D5N and data were recorded on a Bruker AVANCE III of 600 MHz for $^1$H-NMR and 150 MHz for $^{13}$C-NMR, respectively.
