



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 synthetases (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.

© 2016 Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

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.

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-derived 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 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 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,

Table 1
Primers used for RT-PCR analysis.

Primers	Sequence (5'–3')
<i>marRI</i> -F	CGCAGAGTTCGGAGGACGAG
<i>marRI</i> -R	ACCCCCGTAATGCGAAACGAAG
<i>marRII</i> -F	CATGACCCTGCTCCCGAAC
<i>marRII</i> -R	CAGTTCCTTCAACCGGTGCGC
<i>marRIII</i> -F	GACTGGCCACCACCATCGAG
<i>marRIII</i> -R	GAAACGGTCCAGCACGTCGTG
<i>marRIV</i> -F	GAGCTGACCGCTCACTCTTTC
<i>marRIV</i> -R	GGTTGGTTCACGACGCC

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%/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

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			NRPS	Hybrid PKS/NRPS	Non-NRPS siderophore
			Type I	Type II	Type III			
<i>S. marokkonensis</i> M10	7.33	≥26	≥3	–	1	≥6	1	2
<i>S. coelicolor</i> A3(2)	8.72	≈8	2	2	3	3	–	1
<i>S. griseus</i> IFO 13350	8.55	34	–	–	–	–	–	–
<i>S. avermitilis</i>	9.03	38	7	2	2	6	1	1
<i>S. albus</i> J1074	6.84	22	2	–	1	3	3	2
<i>S. tropica</i> CNB-440	5.18	17	1	2	1	3	4	1
<i>N. farcinica</i> IFM 10152	6.01	–	4	1	1	1	7	–

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

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 JX397876. 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^T. It was reported that *S. marokkonensis* Ap1^T was isolated from rhizosphere soil of the indigenous Moroccan plant *Argania spinosa* L.,²³ and a producer of antifungal polyene macrolides.²⁴ 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* J1074,²⁵ 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 *pkc1* 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 *pkc-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.

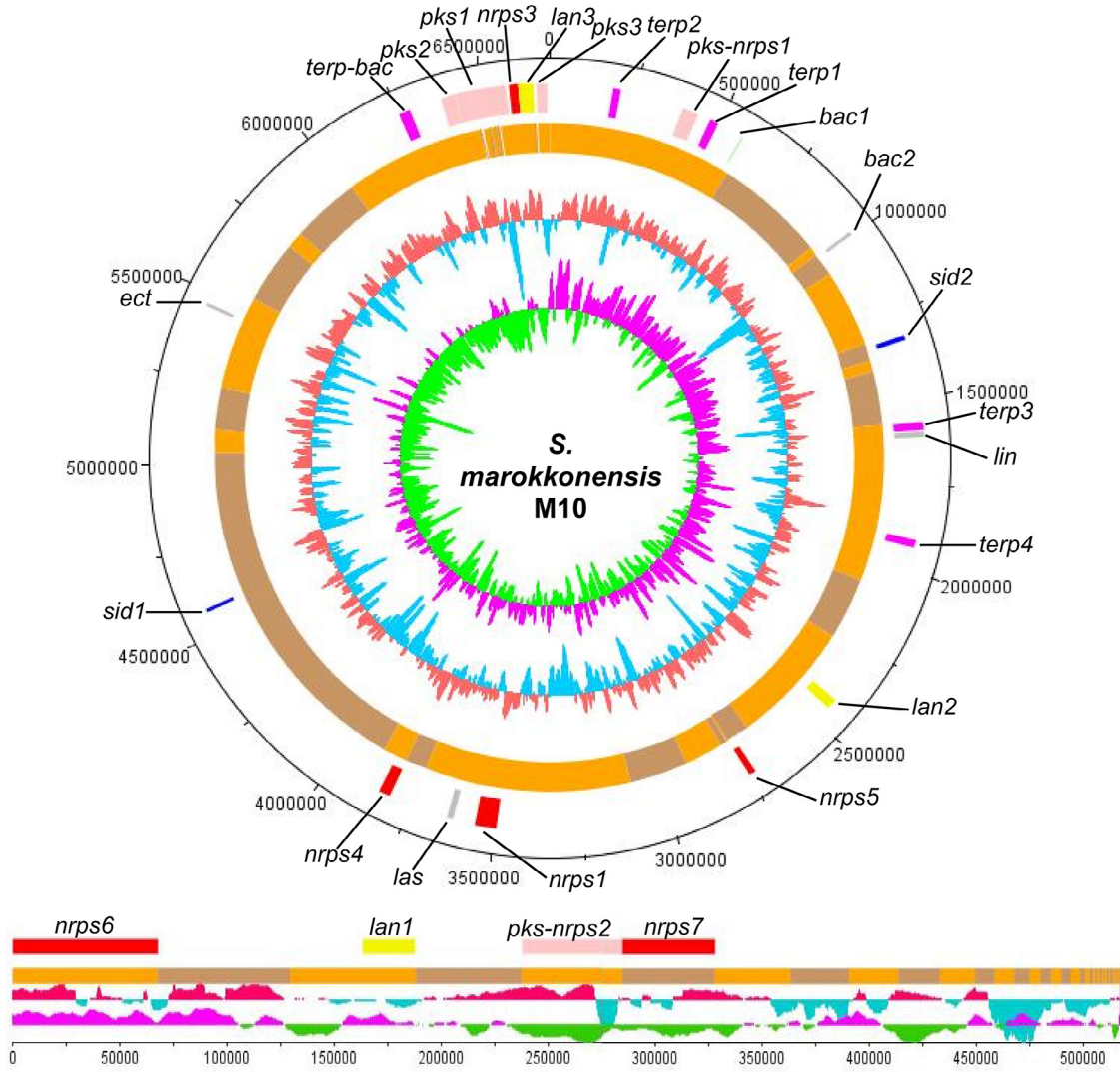
3.2. Biosynthetic gene cluster associated with secondary metabolites

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),³ and several other commonly studied strains, including *S. griseus* IFO 13350,²⁵ *S. avermitilis*,²⁶ *S. albus* J1074,²⁷ *Salinispora tropica* CNB-440,²⁸ and *Nocardia farcinica* IFM 10152.²⁹ It is clear that comparing with these commonly known natural product producers, *S. marokkonensis* M10 has more NRPS gene clusters. Sequence analysis revealed that over 26 gene clusters in *S. marokkonensis* M10 were predicted to be involved in the biosynthesis of multiple secondary metabolites, including PKS, NRPS, siderophores, lantibiotics, terpenoids, and so on.

The putative natural product biosynthetic gene clusters were summarized in Table 3. The *pkc1* cluster which belonged to the type I PKS is the largest biosynthetic gene cluster in the genome of *S. marokkonensis* M10. The *pkc1* gene cluster shares high nucleotide sequence similarity with the candididin/FR008 biosynthetic gene cluster from *Streptomyces* sp. FR-008.^{30,37} Another *pkc* cluster contains multiple genes similar to that in the Herboxidiene biosynthetic gene cluster.³¹ Other natural product related gene clusters were also predicted based on the genome sequence, including six 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,³² one antimycin-like³³ and one desotamides-like³⁴ gene cluster, respectively. We also identified a frotalamides-like³⁵ cluster containing hybrid PKS-NRPS, and one terpenoid cluster similar to germacrene D.³⁶

The deduced functions of proteins encoded by *pkc1* biosynthetic gene cluster were listed in Table 4, and the *pkc1* gene cluster clearly distinguished itself from the candididin gene cluster from *S. griseus* IMRU 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. sp.* FR-008 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

A



B

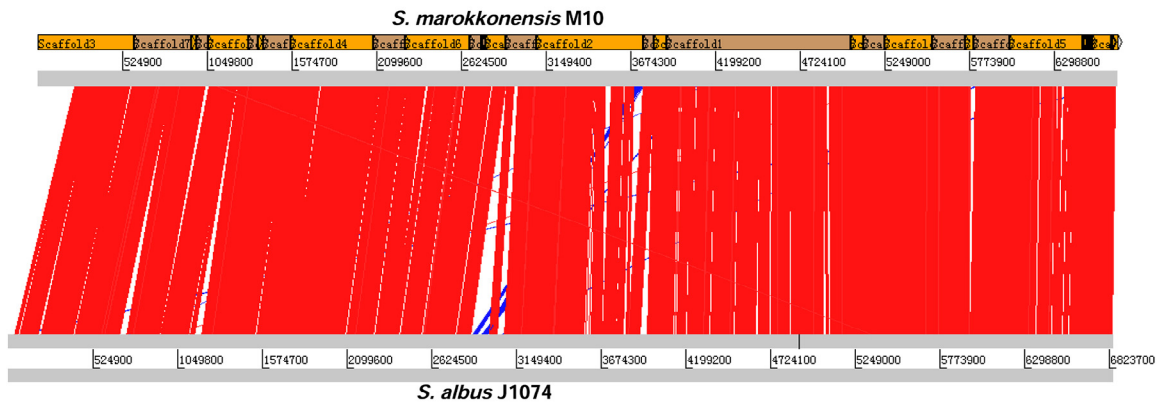


Fig. 1. Linear map of the assembled chromosome of *S. marokkonensis* M10 (illustrated as a circular one) by alignment and the none-assembled scaffolds (illustrated as a liner one) generated from DNAPlotter software. (A) The outer ring and the top line show the locations of natural product-associated 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.

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

No.	Cluster designation	Predicted product	Type	Size (kb)	Gene location	References
1	<i>pks1</i>	FR-008-like polyene macrolides	Type I PKS	147.9	Gene 2271–2287 Scaffold 53, 44, 58, 41, 60, 55, 37 Gene 4799–4824	30
2	<i>pks2</i>	Herboxidiene like	Type III PKS	41.1	Gene 4825–4868	31
3	<i>pks3*</i>	Unknown	Type I PKS	29.5	Gene 4113–4137	
4	<i>pks4#</i>	Unknown	Type I PKS	22.8	Gene 4239–4247	
5	<i>pks5#</i>	Unknown	Type I PKS	19.2	Gene 4256–4268	
6	<i>nrps1</i>	Mannopeptimycin-like hexapeptide	NRPS	62.2	Gene 2638–2677	32
7	<i>nrps2</i>	Unknown	NRPS	61.2	Gene 3871–3916	
8	<i>nrps3</i>	Antimycin like	NRPS	43.4	Gene 2249–2264	33
9	<i>nrps4</i>	Desotamide like	NRPS	58.5	Gene 2337–2356 Gene 3198–3218	34
10	<i>nrps5</i>	Unknown peptide	NRPS	20.0	Gene 3173–3196, 5234, 4269–4285, 2085–2086	
11	<i>nrps6#</i>	Unknown peptide	NRPS	68.2	Gene 3259–3292	
12	<i>nrps7#</i>	Unknown peptide	NRPS	43.4	Gene 4100–4112	
13	<i>pks-nrps1</i>	Frontalamides like	PKS-NRPS	49.4	Gene 3716–3749	35
14	<i>pks-nrps2#</i>	Unknown	PKS-NRPS	47.2	Gene 3557–3579	
15	<i>lan1</i>	Lantipeptide	Lantibiotic	25.3	Gene 3367–3392	
16	<i>lan2</i>	Thiopeptide-lantipeptide	Lantibiotic	32.5	Gene 5459–5488	
17	<i>Lan3*</i>	Lantipeptide-NRPS	Lantipeptide-NRPS	45.7	Gene 2220–2248	
18	<i>terp1</i>	Hopene-like terpene	Terpene	26.6	Gene 3654–3679	
19	<i>terp2</i>	Unknown terpene	Terpene	20.3	Gene 3914–3928	
20	<i>terp3</i>	Germacrene D	Terpene	20.2	Gene 4291–4312	36
21	<i>terp4</i>	Unknown terpene	Terpene	21.2	Gene 4604–4621	
22	<i>terp-bac</i>	Isorenieratene like	Terpene-bacteriocin	31.1	Gene 4961–4955	
23	<i>sid1</i>	Desferrioxamine_B	NRPS-independent siderophore	12.0	Gene 0673–0684	
24	<i>sid2</i>	Unknown siderophore	NRPS-independent siderophore	15.1	Gene 3403–3416	
25	<i>lin</i>	Unknown	Other	17.9	Gene 4319–4335	
26	<i>las</i>	Unknown	Other	20.9	Gene 2547–2563	
27	<i>bac1</i>	Bacteriocin	Other	5.5	Gene 3587–3590	
28	<i>bac2</i>	Bacteriocin	Other	10.9	Gene 3068–3079	
29	<i>ect</i>	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.

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 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%	AAQ82560
PabC	257	ADC lyase	PabC	96%/97%	AAQ82550
MarO	400	FAD-dependent monooxygenase	FscO	99%/98%	AAQ82549
MarA	1744	Type I polyketide synthase	FscA	96%/96%	AAQ82561
MarB	1806	Type I polyketide synthase	FscB	96%/96%	AAQ82565
MarC	11006	Type I polyketide synthase	FscC	-/*	AAQ82564
MarD	9472	Type I polyketide synthase	FscD	97%/97%	AAQ82568
MarE	7848	Type I polyketide synthase	FscE	-/**	AAQ82567
MarF	2040	Type I polyketide synthase	FscF	94%/94%	AAQ82566
MarRI	148	LuxR family transcriptional regulator	FscRI	98%/97%	YP_007749177
MarRII	942	LuxR family transcriptional regulator	FscRII	99%/99%	AAQ82552
MarRIII	1017	LuxR family transcriptional	FscRIII	97%/98%	AAQ82553
MarRIV	963	LuxR family transcriptional	FscRIV	97%/98%	AAQ82554
MarMI	458	Glycosyltransferase	FscMI	99%/99%	YP_007749173
MarMII	352	GDP-ketosugar aminotransferase	FscMII	99%/99%	AAQ82556
MarMIII	349	GDP-mannose-4, 6-dehydratase	FscMIII	99%/99%	AAQ82569
MarP	393	Cytochrome P450 monooxygenase	PimG	99%/99%	CAC20928
MarFE	64	Ferredoxin	FscFE	100%/100%	YP_007749170
MarTE	256	Type II thioesterase	FscTE	98%/98%	AAQ82559
MarTI	335	ABC transporter	FscTI	97%/98%	YP_007749166
MarTII	262	ABC transporter	FscTII	90%/95%	AAQ82563

* Query cover: 88%.

** Query cover: 78%.

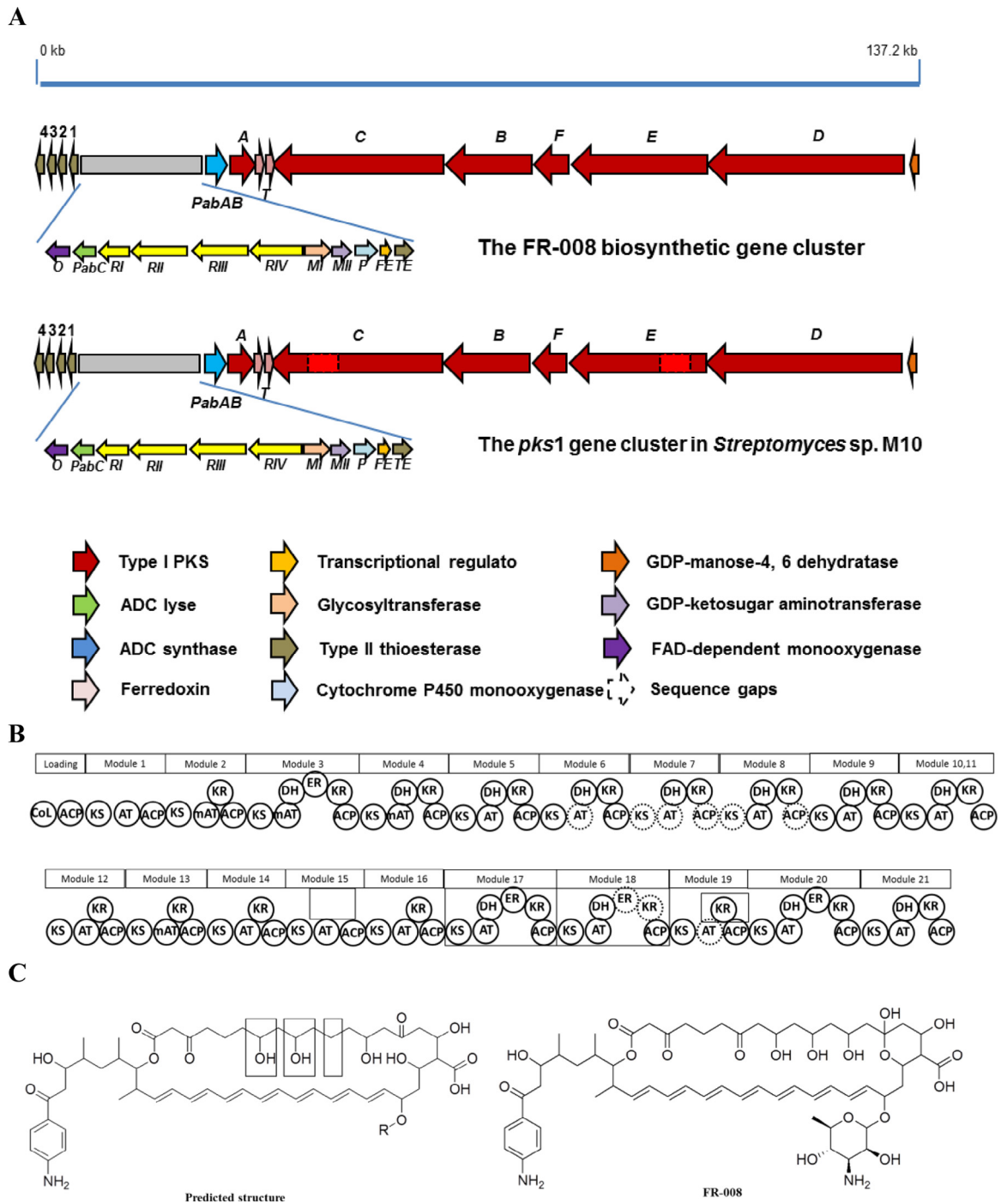


Fig. 2. The polyene-related *pks1* gene cluster in *S. marokkonensis* M10. (A) Comparison of the gene cluster with the known FR-008 gene cluster. (B) The predicted biosynthetic pathway to synthesize the backbone of the marcolides. (C) The predicted product was compared with that of FR-008. The boxes indicate the putative differences.

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 *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 *marRI*, *marRII*, *marRIII* and *marRIV* were investigated using the RNA samples collected 2, 4 and 6 days post-inoculation. As illustrated in Fig. 3, *marRII* and *marRIII* were expressed in the 2nd day but *marRI*

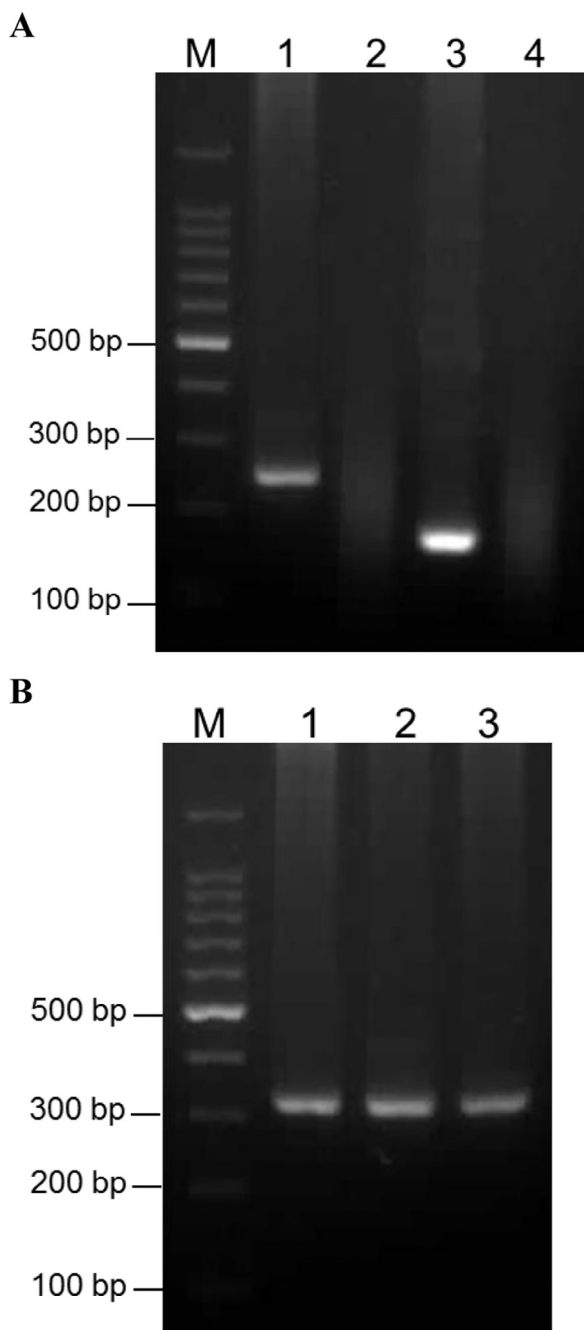


Fig. 3. Expression analysis of the four regulatory genes involved in *pkS1* biosynthetic gene cluster. (A) PCR products of *marRII* and *marRIII*. M, 100 bp DNA ladder marker; 1, *marRII*, total RNA extracted in 2nd day; 2, *marRII*, total RNA extracted in 4th day; 3, *marRIII*, total RNA extracted in 2nd day; 4, *marRIII*, total RNA extracted in 4th day. (B) Positive control by Hrd B. M, 100 bp DNA ladder marker; 1, 2 and 3 represented the total RNA extracted in 2nd, 4th and 6th day, respectively.

and *marRIV* didn't show any expression in the three selected time points, implying that *marRII* and *marRIII* 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 polyene compounds were further exemplified by the isolation of compound 9-04 from PF1. In the $^1\text{H-NMR}$ spectrum of 9-04, the peaks at $\delta 5.3$ – 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 $^{13}\text{C-NMR}$ spectrum of 9-04 at $\delta 176$ 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 $^1\text{H-NMR}$ 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 streptomycete *S. marokkonensis* M10, which is a similar strain of *S. marokkonensis* Ap1^T (99.02% 16S rDNA sequence similarity). The type strain *S. marokkonensis* Ap1^T 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 *fscRII*, *fscRIII* and partial *fscRI* led to the absence of FR-008.³⁰ The regulatory genes (*marRI*, *marRII*, *marRIII* and *marRIV*) were deduced to act as the same roles in *pkS1* gene cluster. The expression of *marRII* and *marRIII* 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 *marRII* and *marRIII* 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

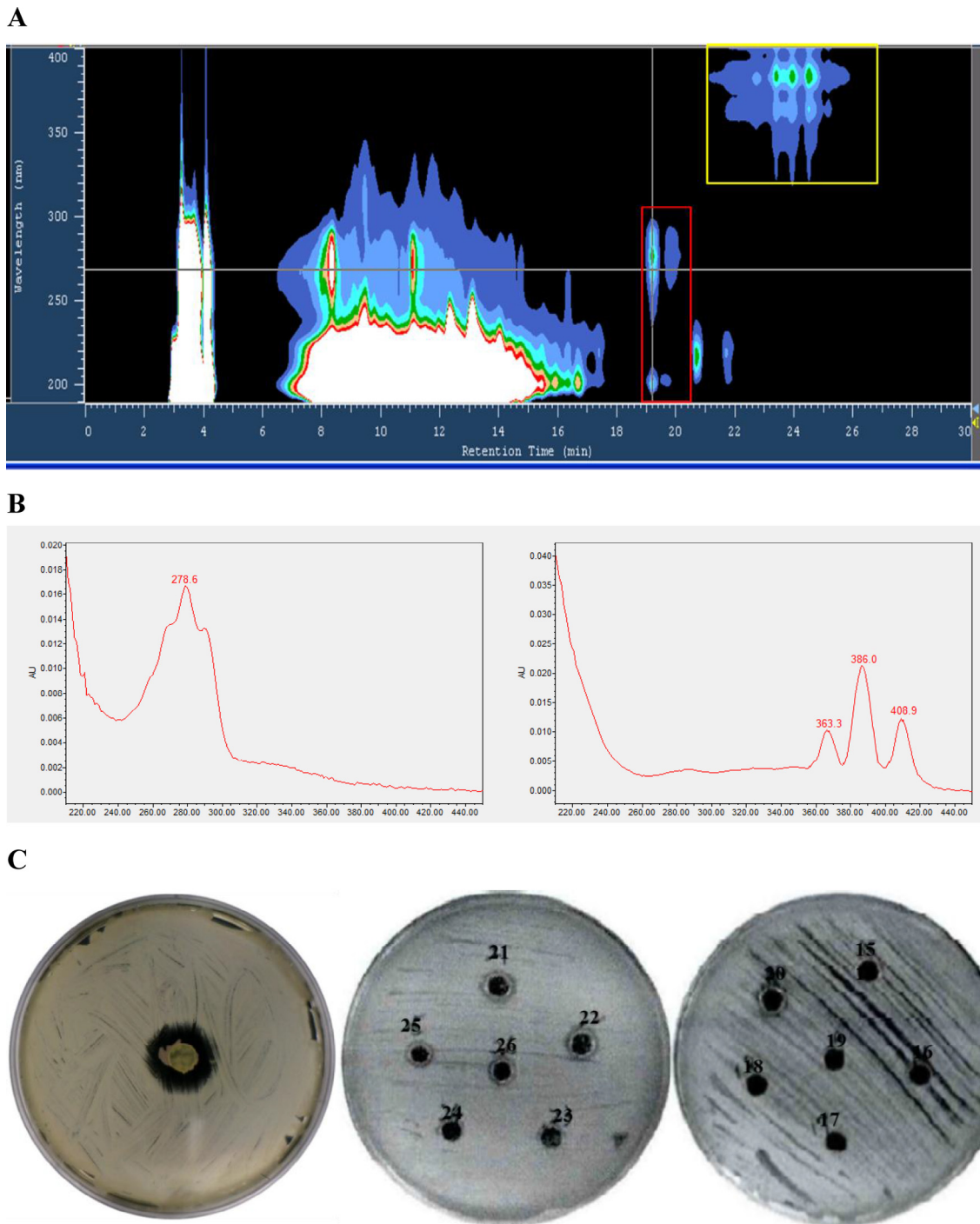
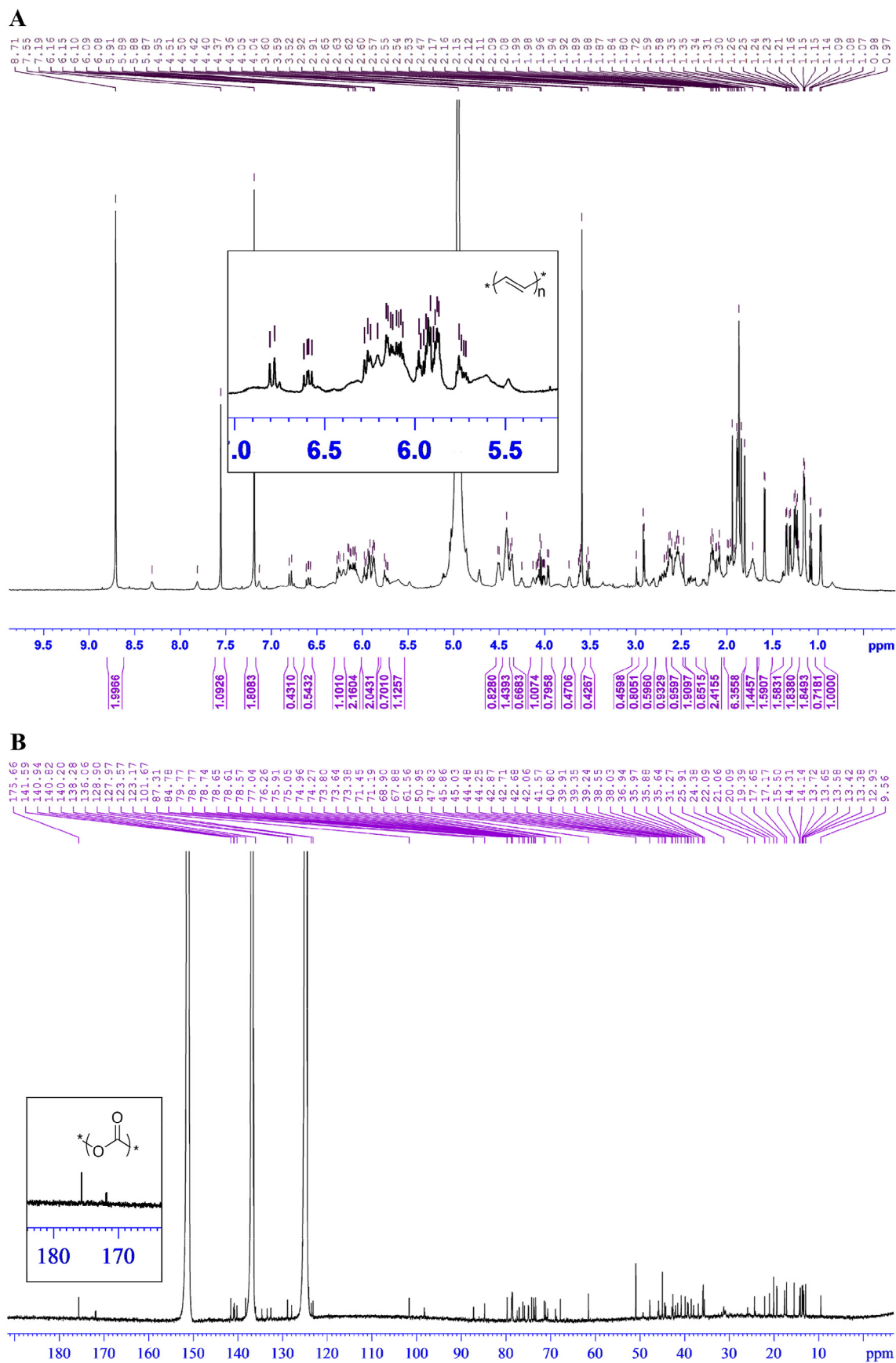


Fig. 4. Identification of the polyene families in the BuOH layer crude extract of *S. marokkonensis* M10 on A1 agar plate. (A) HPLC profile of polyene families PF1 (red-boxed) and PF2 (yellow-boxed). (B) Characteristic UV adsorption spectrum of PF1 and PF2. (C) The antifungal activity against *C. albicans* 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.

products, with unrivaled track record as antifungal antibiotics, were initially recognized as amphotericin B and nystatin,^{38,39} and subsequently candicidin, as well as a series of newly isolated polyene compounds such as bahamaolides, marinisporolides, wortmannilactones and takanawaenes^{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 ¹H-NMR also verified the analysis above and more specifically, 9-04 was identified as triene macrolides in consideration of the lactone deduced by ¹³C-NMR spectrum.

Although the precise organization of *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 *pks1* in *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



cluster is indeed related to the biosynthesis of PF1 and PF2 mentioned above.

The polyene natural products via putative biosynthetic gene cluster mining from a previously unsequenced *Streptomyces* species highlighted the strength of genome mining which effectively bridged the gap between chemotypes and genotypes. With thousands of bacterial strains being sequenced, we believed that genome mining would provide a comprehensive understanding of interested microorganism and could be powerful tool for the discovery of novel natural products.

Acknowledgements

The authors are grateful to Wence Jiao for his assistance on the studies of polyene compounds.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk HP, et al. Taxonomy, physiology, and natural products of actinobacteria. *Microbiol Mol Biol Rev* 2015;**80**:1–43.
- Manivasagan P, Kang KH, Sivakumar K, Li-Chan EC, Oh HM, Kim SK. Marine actinobacteria: an important source of bioactive natural products. *Environ Toxicol Pharmacol* 2014;**38**:172–88.
- Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 2002;**417**:141–7.
- Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A* 2001;**98**:12215–20.
- Katz L, Baltz RH. Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* 2016;**43**:155–76.
- Li X, Lei X, Zhang C, Jiang Z, Shi Y, Wang S, et al. Complete genome sequence of *Streptomyces globosporus* C-1027, the producer of an enediyne antibiotic lidamycin. *J Biotechnol* 2016;**S0168-1656**(16):pii.
- Xu MJ, Wang JH, Bu XL, Yu HL, Li P, Ou HY, et al. Deciphering the streamlined genome of *Streptomyces xiamenensis* 318 as the producer of the anti-fibrotic drug candidate xiamenmycin. *Sci Rep* 2016;**6**:18977.
- Song JY, Yoo YJ, Lim SK, Cha SH, Kim JE, Roe JH, et al. Complete genome sequence of *Streptomyces venezuelae* ATCC 15439, a promising cell factory for production of secondary metabolites. *J Biotechnol* 2016;**219**:57–8.
- Scheffler RJ, Colmer S, Tynan H, Demain AL, Gullo VP. Antimicrobials, drug discovery, and genome mining. *Appl Microbiol Biotechnol* 2013;**97**:969–78.
- Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. *Chembiochem* 2009;**10**:625–33.
- Zhao XQ. Genome-based studies of marine microorganisms to maximize the diversity of natural products discovery for medical treatments. *Evid Based Complement Alternat Med* 2011;**384572**:2011.
- Corre C, Challis GL. New natural product biosynthetic chemistry discovered by genome mining. *Nat Prod Rep* 2009;**26**:977–86.
- Nett M, Ikeda H, Moore BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* 2009;**26**:1362–84.
- Zhao XQ, Li WJ, Jiao WC, Li Y, Yuan WJ, Zhang YQ, et al. *Streptomyces xinghaiensis* sp. nov., a new marine actinomycete derived from a marine sediment sample. *Int J Syst Evol Microbiol* 2009;**59**:2870–4.
- Chen C, Feng WW, Qin S, Zhao XQ. *Streptomyces xiaopingdaonensis* sp. nov., a novel marine actinomycete isolated from the sediment of Xiaopingdao in Dalian, China. *Antonie Van Leeuwenhoek* 2015;**107**:511–8.
- Zhao XQ, Yang TH. Draft genome sequence of the marine sediment-derived actinomycete *Streptomyces xinghaiensis* NRRL B24674T. *J Bacteriol* 2011;**193**:5543.
- Zhao XQ, Geng X, Chen C, Chen LY, Jiao WC, Yang C. Draft genome sequence of the marine actinomycete *Streptomyces sulphureus* L180, isolated from marine sediment. *J Bacteriol* 2012;**194**:4482.
- Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* 2012;**62**:716–21.
- Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics* 2008;**24**:713–4.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119.
- Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 2008;**24**:2672–6.
- Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 2011;**39**:W339–46.
- Bouizgarne B, Lanoot B, Loqman S, Spröer C, Klenk HP, Swings J, et al. *Streptomyces marokkonensis* sp. nov., isolated from rhizosphere soil of *Argania spinosa* L. *Int J Syst Evol Microbiol* 2009;**59**:2857–63.
- Bouizgarne B, Hadrami IE, Ouhdouch Y. Novel production of isochainin by a strain of *Streptomyces* sp. isolated from rhizosphere soil of the indigenous Moroccan plant *Argania spinosa* L. *World J Microb Biot* 2006;**22**:423–9.
- Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, et al. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* 2008;**190**:4050–60.
- Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A* 2001;**98**:12215–20.
- Zaburannyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics* 2014;**15**:97.
- Udwary DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, et al. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci U S A* 2007;**104**:10376–81.
- Ishikawa J, Yamashita A, Mikami Y, Hoshino Y, Kurita H, Hotta K, et al. The complete genomic sequence of *Nocardia farcinica* IFM 10152. *Proc Natl Acad Sci U S A* 2004;**101**:14925–30.
- Chen S, Huang X, Zhou X, Bai L, He J, Jeong KJ, et al. Organizational and mutational analysis of a complete FR-008/candidicin gene cluster encoding a structurally related polyene complex. *Chem Biol* 2003;**10**:1065–76.
- Pokhrel AR, Dhakal D, Jha AK, Sohng JK. Herboxidiene biosynthesis, production, and structural modifications: prospect for hybrids with related polyketide. *Appl Microbiol Biotechnol* 2015;**99**:8351–62.
- Magarvey NA, Haltli B, He M, Greenstein M, Hucul JA. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob Agents Chemother* 2006;**50**:2167–77.
- Yan Y, Zhang L, Ito T, Qu X, Asakawa Y, Awakawa T, et al. Biosynthetic pathway for high structural diversity of a common dilactone core in antimycin production. *Org Lett* 2012;**14**:4142–5.
- Li Q, Song Y, Qin X, Zhang X, Sun A, Ju J. Identification of the biosynthetic gene cluster for the anti-infective desotamides and production of a new analogue in a heterologous host. *J Nat Prod* 2015;**78**:944–8.
- Blodgett JA, Oh DC, Cao S, Currie CR, Kolter R, Clardy J. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc Natl Acad Sci U S A* 2010;**107**:11692–7.
- Lücker J, Bowen P, Bohlmann J. Vitis vinifera terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (–)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry* 2004;**65**:2649–59.
- Campelo AB, Gil JA. The candidicin gene cluster from *Streptomyces griseus* IMRU 3570. *Microbiology* 2002;**148**(1):51–9.
- Oura M, Sternberg TH, Wright ET. A new antifungal antibiotic, amphotericin B. *Antibiot Annu* 1955;**3**:566–73.
- Dutcher J, Walters D, Wintersteiner O. Nystatin. III. Mycosamine: preparation and determination of structure. *J Org Chem* 1963;**28**:995–9.
- Dong Y, Yang J, Zhang H, Lin J, Ren X, Liu M, et al. Wortmannilactones A–D, 22-membered triene macrolides from the marine actinomycete *wortmannii*. *J Nat* 2006;**69**:128–30.
- Fukuda T, Kim YP, Iizima K, Fukuda T, Matsumoto A, Takahashi Y, et al. Takanawaenes, novel antifungal antibiotics produced by *Streptomyces* sp. K99-5278. II. Structure elucidation. *J Antibiot (Tokyo)* 2003;**56**:454–8.
- Kim DG, Moon K, Kim SH, Park SH, Park S, Lee SK, et al. Bahamaolides A and B, antifungal polyene polyol macrolides from the marine actinomycete *Streptomyces* sp. *J Nat* 2012;**75**:959–67.
- Kwon HC, Kauffman CA, Jensen PR, Fenical W. Marinisporolides, polyene-polyol macrolides from a marine actinomycete of the new genus *Marinispora*. *J Org Chem* 2009;**74**:675–84.