

## RESEARCH ARTICLE

# Complete genome sequence of the deep South China Sea-derived *Streptomyces niveus* SCSIO 3406, the producer of cytotoxic and antibacterial marfuraquinocins

Qinghua Zhu<sup>1</sup>, Weige Cheng<sup>2</sup>, Yongxiang Song<sup>2</sup>, Qing He<sup>1</sup>, Jianhua Ju<sup>2</sup>\*, Qinglian Li<sup>2</sup>\*

**1** College of Life Science, Dezhou University, Dezhou, China, **2** CAS Key Laboratory of Tropical Marine Bio-Resources and Ecology, Guangdong Key Laboratory of Marine Materia Medica, RNAM Center for Marine Microbiology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

✉ These authors contributed equally to this work.

\* [liql@scsio.ac.cn](mailto:liql@scsio.ac.cn) (QL); [jjj@scsio.ac.cn](mailto:jjj@scsio.ac.cn) (JJ)



## OPEN ACCESS

**Citation:** Zhu Q, Cheng W, Song Y, He Q, Ju J, Li Q (2021) Complete genome sequence of the deep South China Sea-derived *Streptomyces niveus* SCSIO 3406, the producer of cytotoxic and antibacterial marfuraquinocins. PLoS ONE 16(3): e0248404. <https://doi.org/10.1371/journal.pone.0248404>

**Editor:** Luis Angel Maldonado Manjarrez, Universidad Nacional Autonoma de Mexico Facultad de Quimica, MEXICO

**Received:** January 30, 2020

**Accepted:** February 25, 2021

**Published:** March 23, 2021

**Copyright:** © 2021 Zhu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** *S. niveus* SCSIO 3406 has been deposited in the China General Microbiological Culture Collection Center under the number of CGMCC 7863 and its complete genome sequence has been deposited in GenBank under accession number CP018047. The BioProject designation for this project is PRJNA328345.

**Funding:** This work was supported by the National Natural Science Foundation of China (41306140)

## Abstract

*Streptomyces niveus* SCSIO 3406 was isolated from a sediment sample collected from South China Sea at a depth of 3536 m. Four new sesquiterpenoid naphthoquinones, marfuraquinocins A-D, and two new geranylated phenazines, i. e. phenaziterpenes A and B, were isolated from the fermentation broth of the strain. Here, we present its genome sequence, which contains 7,990,492 bp with a G+C content of 70.46% and harbors 7088 protein-encoding genes. The genome sequence analysis revealed the presence of a 28,787 bp gene cluster encoding for 24 open reading frames including 1,3,6,8-tetrahydroxynaphthalene synthase and monooxygenase, seven phenazine biosynthesis proteins, two prenyl-transferases and a squalene-hopene cyclase. These genes are known to be necessary for the biosynthesis of both marfuraquinocins and phenaziterpenes. Outside the gene cluster (and scattered around the genome), there are seven genes belonging to the methylerythritol phosphate pathway for the biosynthesis of the essential primary metabolite, isopentenyl diphosphate, as well as six geranyl diphosphate/farnesyl diphosphate synthase genes. The strain *S. niveus* SCSIO 3406 showed type I PKS, type III PKS and nonribosomal peptide synthetase cluster. The sequence will provide the genetic basis for better understanding of biosynthesis mechanism of the above mentioned six compounds and for the construction of improved strain for the industrial production of antimicrobial agents.

## Introduction

Deep-sea *Streptomyces* are widely recognized as an emerging source of novel and bioactive secondary metabolites [1]. They have been phylogenetically classified in 13 groups (MAR1--MAR13) [2]. The MAR4 group is a rich source of polyketide-terpenoid secondary metabolites, such as marinone, azamerone, and napyradiomycins [3–5]. As a member of the MAR4 group, *Streptomyces* sp. CNQ-509 could produce two polyketide-terpenoids (naphterpin and

received by QZ, the Pearl River S&T Nova Program of Guangzhou (201806010109), the Special Support Program for Training High-Level Talents in Guangdong (2016TQ03R288) received by QL and the Program of Guangzhou Science and Technology Plan (201707010454) received by YS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

debromomarinone), five new farnesyl- $\alpha$ -nitropyrroles nitropyrrolins A–E and O-prenylated phenazines marinophenazine A–B [6]. Here, *S. niveus* SCSIO 3406 was isolated from a South China Sea sediment sample collected at a depth of 3536 m. Four new sesquiterpenoid naphthoquinones, marfuraquinocins A–D (1–4) (Fig 1), which exhibited antibacterial activities against *Staphylococcus aureus* ATCC 29213 or methicillin-resistant *Staphylococcus epidermidis* (MRSE) were previously isolated from the strain SCSIO 3406 [7]. Additionally, two new geranylated phenazines, phenaziterpenes A and B (5–6) (Fig 1), were also isolated from this strain in spite of low production [7]. In order to gain insights about the genetic basis of the above six compounds and about the discovery of further new natural products, the genome of *S. niveus* SCSIO 3406 was sequenced.

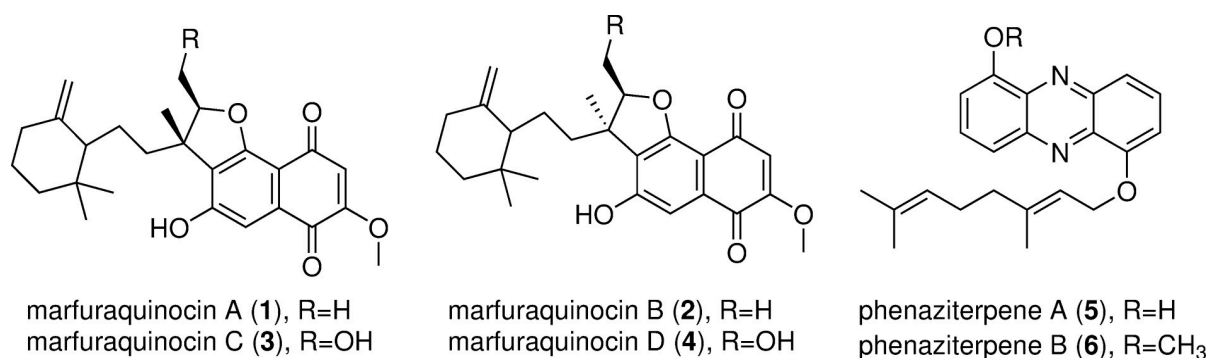
## Materials and methods

*S. niveus* SCSIO 3406 was cultivated in trypticase soy broth and grown for 2 days at 28 °C with 200-rpm aeration. High-molecular-weight DNA was prepared using standard genomic DNA isolation method [8].

Genomic DNA was then used to generate two libraries, one is paired-end (PE) library with insert sizes of 300~500 bp and the other is a SMRTbell<sup>TM</sup> template library with insert sizes of about 8~10 kb. The complete genome of *S. niveus* SCSIO 3406 was subsequently sequenced using a combination of PacBio RSII sequencing (Pacific Biosciences) and Illumina Hiseq 2500 technologies at Biozeron Biotech Co., LTD (Shanghai, China). Genome assembly was *de novo* performed with SOAPdenovo v2.04 [9] and Celera Assembler 8.0 [10]. Putative protein-coding sequences were predicted by Glimmer 3.02 [11]. Gene functional annotation was performed using BLASTP with Nr, String, COG and KEGG databases. rRNA, tRNA were predicted using RNAmmer v1.2 and NCBI Prokaryotic Genome Annotation Pipeline respectively. Protein coding genes were analyzed for COG functional annotation using WebMGA server [12]. CRISPRFinder, freely accessible at <http://crispr.i2bc.paris-saclay.fr> is used to find clustered regularly interspaced short palindromic repeats (CRISPRs) in *S. niveus* SCSIO 3406 genome. Genes involved in secondary metabolic pathways were predicted using antiSMASH 2.0 (<http://antismash.secondarymetabolites.org/>) [13].

## Results and discussion

The complete genome of the strain *S. niveus* SCSIO 3406 consisted of a single linear chromosome of 7,990,492 bp with an average GC content of 70.46% without plasmids. Annotation revealed a total of 7,088 protein-coding genes, 6 rRNA operons with the order 5S-16S-23S, and 65 tRNA genes which could transfer all the twenty proteinic amino acids (Table 1, Fig 2). Of



**Fig 1. The chemical structure of compounds marfuraquinocins A–D (1–4) and phenaziterpenes A–B (5–6).**

<https://doi.org/10.1371/journal.pone.0248404.g001>

**Table 1. Genomic features of *S. niveus* SCSIO 3406.**

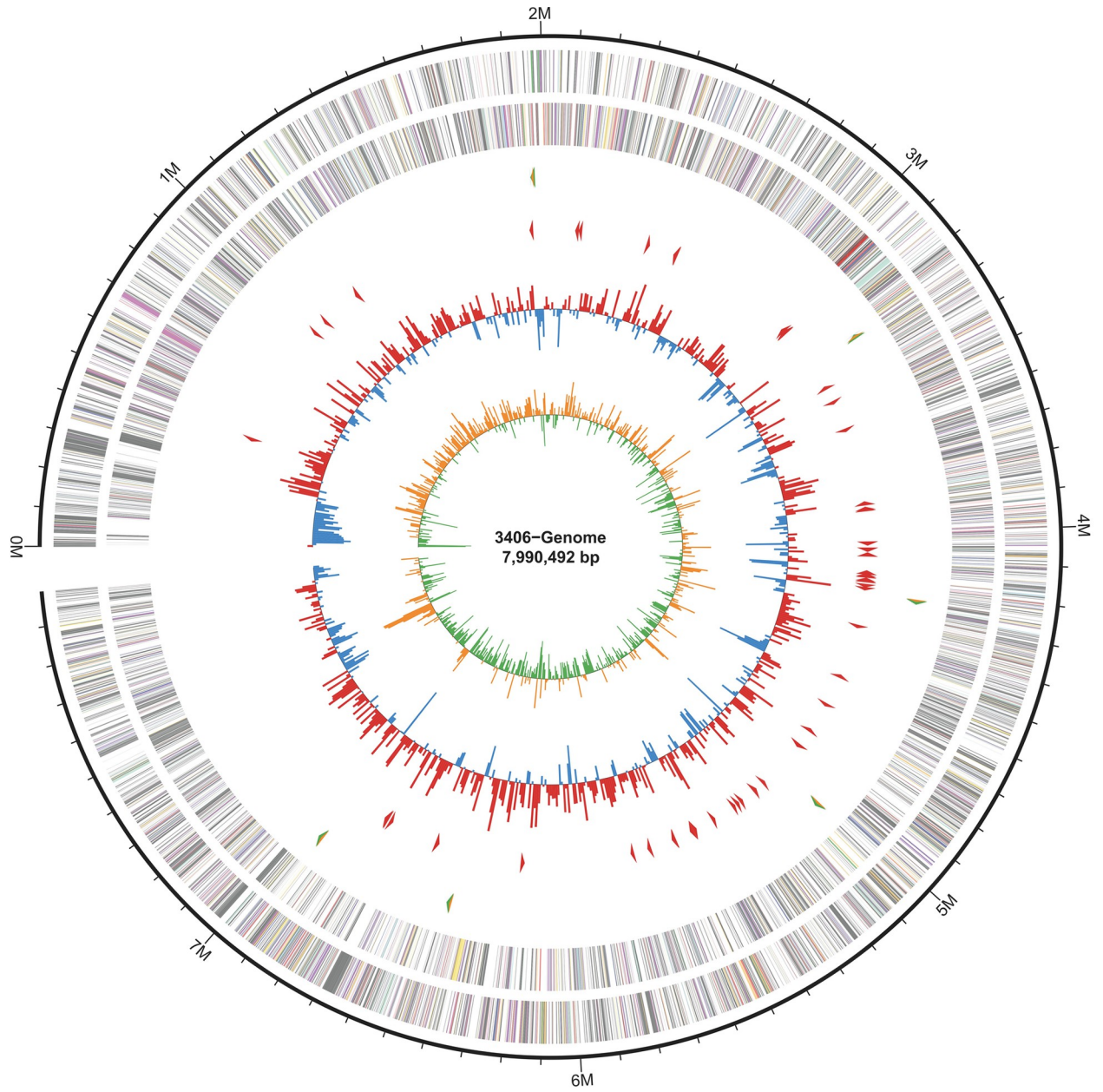
Features	Values
genome topology	linear
genome size (bp)	7,990,492
the sequencing coverage	315.5
average G+C content (%)	70.46
protein-coding genes	7088
rRNA genes	6, 6, 6 (5S, 16S, 23S)
tRNA genes	65

<https://doi.org/10.1371/journal.pone.0248404.t001>

the 7,088 protein-coding genes, 3345 (38.3%) were classified into different COG functional categories (Table 2, Fig 2), 4157 shared significant homology with genes from *S. niveus* NCIMB 11891, the producer of the gyrase inhibitor novobiocin [14]. Most of the genes in *S. niveus* SCSIO 3406 were associated with functions such as “transcription” (K), “amino acid transport and metabolism” (E), “carbohydrate transport and metabolism” (G), “Energy production and conversion” (C), and in particular, “Secondary metabolites biosynthesis, transport and catabolism” (Q) (Table 2).

Both marfuraquinocins A-D and phenaziterpenes A-B have a moiety of terpene. All terpenoids are synthesized from two precursors: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Two distinct biosynthetic pathways produce the essential primary metabolites IPP and DMAPP: the 2-C-methylerythritol 4-phosphate (MEP) pathway and the mevalonate pathway (MVA) [15]. IPP isomerase catalyzes the interconversion of IPP and DMAPP [16]. Based on the Nr annotation, no unigenes were identified as the genes of MVA pathway. However, seven genes scattered around the genome were identified as being involved in the MEP pathway: *orf00197* coding for 1-deoxy-D-xylulose-5-phosphate synthase, *orf01743* coding for 1-deoxy-D-xylulose-5-phosphate reductoisomerase, *orf03854* coding for 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, *orf04260* coding for 4-diphosphocytidyl-2C-methyl-D-erythritol kinase, *orf03855* coding for 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *orf04113* coding for 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, and *orf04123* coding for 4-hydroxy-3-methylbut-2-enyl diphosphate reductase. This implies that the isoprenoid moieties of marfuraquinocins A-D and phenaziterpenes A-B are derived from MEP.

We identified gene *orf01170* coding for type III polyketide synthase and *orf01166* for 1,3,6,8-tetrahydroxynaphthalene (THN) monooxygenase, responsible for the formation and modification of THN successively [17] to generate THN moiety of marfuraquinocins A-D. In many cases, the encoding genes responsible for the antibiotic biosynthesis are clustered in a continuous genomic DNA region, and usually in association with one or more genes that regulate their transcription and with resistance genes [18]. Therefore, the other biosynthetic genes for marfuraquinocins A-D are expected to exist in the vicinity of *orf01170* and *orf01166*. By bioinformatic analysis of the genes located upstream and downstream of *orf01170* and *orf01166*, a ~28.7 kb continuous DNA segment encoding for 24 open reading frames (ORFs) was predicted to contain the putative gene cluster of marfuraquinocins A-D. The function of the individual ORFs was deduced by BLAST analysis. The results are summarized in Table 3 and Fig 3. The identified DNA segment also included a squalene-hopene cyclase gene (*orf01153*), which is predicted to catalyze the cyclization of sesquiterpene to generate one of the parent skeletons of marfuraquinocins A-D. In addition, we found that, as shown in Fig 3 and Table 3, this continuous DNA segment also contained orthologues of most genes required for phenazine biosynthesis in two discontinuous loci (*orf01155-01158* and *orf01161-orf01163*).



Non-coding RNA

- tRNA
- 5S RNA
- 16S rRNA
- 23S rRNA

COG Functional categories

- RNA processing and modification
- Chromatin structure and dynamics
- Energy production and conversion
- Cell cycle control, cell division, chromosome partitioning
- Amino acid transport and metabolism
- Nucleotide transport and metabolism
- Carbohydrate transport and metabolism
- Coenzyme transport and metabolism
- Lipid transport and metabolism
- Translation, ribosomal structure and biogenesis
- Transcription
- Replication, recombination and repair
- Cell wall/membrane/envelope biogenesis
- Cell motility
- Posttranslational modification, protein turnover, chaperones
- Inorganic ion transport and metabolism
- Secondary metabolites biosynthesis, transport and catabolism
- General function prediction only
- Function unknown
- Signal transduction mechanisms
- Intracellular trafficking, secretion, and vesicular transport
- Defense mechanisms
- Extracellular structures
- Nuclear structure
- Cytoskeleton
- No assigned COG

**Fig 2. Genome map of *S. niveus* SCSIO 3406.** The genome map consists of six circles. From outside to inside: (1) the outer two circles show COG functional categories of coding regions in both the forward and reverse strands; (2) the third circle shows the location of rRNA genes; (3) the fourth circle shows the location of tRNA genes; (4) the fifth circle shows GC content values; (5) the central circle shows GC-skew values (G–C/G+C) of the third bases of codons measured over the genome.

<https://doi.org/10.1371/journal.pone.0248404.g002>

Although the orthologue of *phzF* necessary for phenazine biosynthesis is absent in this DNA segment, two *phzF* orthologues (*orf05123* and *orf05140*) were identified outside of the gene cluster.

Two prenyltransferase genes (*orf01154* and *orf01164*) were identified in this DNA segment. Prenyltransferases are a class of enzymes that transfer allylic prenyl groups to acceptor molecules [19]. Therefore, *orf01154* and *orf01164* were predicted to be responsible for the condensation reaction between the sesquiterpene moiety and the THN moiety to form the backbone of marfuraquinocins A-D and/or the condensation reaction between the monoterpene moiety and the phenazine moiety to form the backbone of phenaziterpenes A-B. Based on above all, we speculate that the continuous DNA segment of ~28.7 kb was involved in the biosynthetic pathway of both marfuraquinocins A-D and phenaziterpenes A-B.

Outside the gene cluster, we also identified gene *orf00932* encoding isopentenyl pyrophosphate isomerase (IPP isomerase). In particular, six geranyl diphosphate (GDP) synthase and farnesyl diphosphate (FDP) synthase genes (*orf00192*, *orf00915*, *orf01258*, *orf02165*, *orf04121* and *orf06133*) are located outside the ~28.7 kb gene cluster. GDP synthase and FDP synthase catalyze the addition of one and two molecules of IPP to DMAPP, yielding GDP and FDP, respectively [20]. Therefore, these six GDP/FDP synthases are proposed to be responsible for the formation of terpene core moieties of both marfuraquinocins A-D and phenaziterpenes A-B. It is obvious that the genes of the biosynthesis of marfuraquinocins A-D and phenaziterpenes A-B are not clustered at a single locus of the genome.

**Table 2. COG functional categories of the complete genome sequences of *S. niveus* SCSIO 3406.**

COG code	Description	Number of unigenes
B	Chromatin structure and dynamics	1
J	Translation, ribosomal structure and biogenesis	129
K	Transcription	373
L	Replication, recombination and repair	129
D	Cell cycle control, cell division, chromosome partitioning	17
M	Cell wall/membrane/envelope biogenesis	152
O	Posttranslational modification, protein turnover, chaperones	86
T	Signal transduction mechanisms	193
U	Intracellular trafficking, secretion, and vesicular transport	20
V	Defense mechanisms	86
C	Energy production and conversion	195
E	Amino acid transport and metabolism	349
F	Nucleotide transport and metabolism	56
G	Carbohydrate transport and metabolism	345
H	Coenzyme transport and metabolism	108
I	Lipid transport and metabolism	134
P	Inorganic ion transport and metabolism	186
Q	Secondary metabolites biosynthesis, transport and catabolism	118
R	General function prediction only	489
S	Function unknown	179

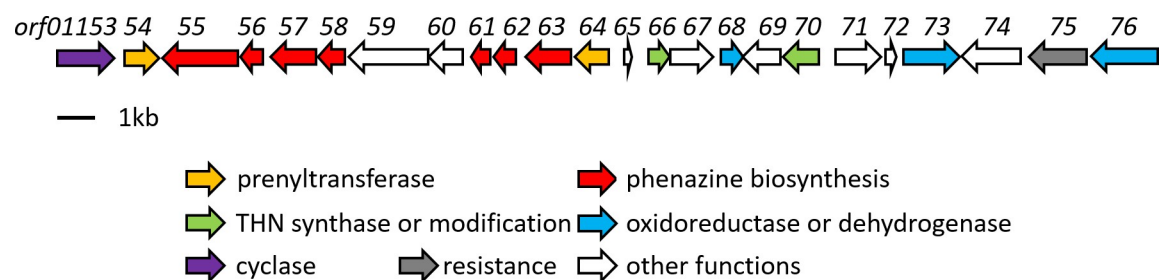
<https://doi.org/10.1371/journal.pone.0248404.t002>

**Table 3. Deduced function of the open reading frames in Fig 3.**

ORFs	Amino acids	Proposed function	Sequence similarity (protein accession number, origin, similarity/identity)
orf01153	498	squalene-hopene/tetraprenyl-beta- curcumene cyclase	SFG55119.1, <i>Streptomyces mirabilis</i> , 53/42
orf01154	305	ABBA prenyltransferase	AFS18550.1, <i>Streptomyces tendae</i> , 77/60
orf01155	657	anthranilate synthase, 2-amino-2- desoxy-isochorismate synthase, PhzE	CDH35391.1, <i>Streptomyces iakyrus</i> , 87/83
orf01156	207	isochorismatase, also known as 2, 3 dihydro-2, 3 dihydroxybenzoate synthase, PhzD	WP_033308464.1, <i>Streptomyces iakyrus</i> , 89/83
orf01157	390	3-deoxy-D-arabino-heptulosonic acid 7- phosphate synthase, PhzC	CDH35389.1, <i>Streptomyces iakyrus</i> , 88/83
orf01158	234	phenazine biosynthesis protein A/B	CDH35388.1, <i>Streptomyces iakyrus</i> , 94/88
orf01159	704	membrane protein	CDH35387.1, <i>Streptomyces iakyrus</i> , 90/83
orf01160	294	RNA polymerase sigma factor	SER87811.1, <i>Streptomyces</i> sp. Yr375, 86/75
orf01161	168	phenazine biosynthesis protein A/B	CDH35386.1, <i>Streptomyces iakyrus</i> , 92/86
orf01162	216	FMN-dependent oxidase, PhzG	CDH35385.1, <i>Streptomyces iakyrus</i> , 89/84
orf01163	408	salicylate hydroxylase, PhzS	CDH35383.1, <i>Streptomyces iakyrus</i> , 85/79
orf01164	301	aromatic prenyltransferase	CDH35382.1, <i>Streptomyces iakyrus</i> , 84/70
orf01165	68	hypothetical protein	WP_062209276.1, <i>Streptomyces</i> sp. NBRC 109706, 78/65
orf01166	185	1,3,6,8-tetrahydroxynaphthalene monooxygenase	SDJ65950.1, <i>Actinopolyspora mzabensis</i> , 84/74
orf01167	380	pyridoxal phosphate-dependent aminotransferase	WP_049565889.1, <i>Streptomyces</i> sp. SBT349, 93/87
orf01168	203	NAD(P)H:quinone oxidoreductase	WP_079314593.1, <i>Microbispora</i> sp. GKU 823, 81/71
orf01169	332	SAM-dependent methyltransferase	WP_073501777.1, <i>Streptomyces paucisporeus</i> , 85/76
orf01170	317	type III polyketide synthase	SHN18252.1, <i>Streptomyces paucisporeus</i> , 90/83
orf01171	404	putative transcriptional regulator	CDH35377.1, <i>Streptomyces iakyrus</i> , 77/71
orf01172	100	hypothetical protein	WP_047018058.1, <i>Streptomyces</i> sp. CNQ-509, 74/57
orf01173	505	FAD/FMN-containing dehydrogenase	SFF48976.1, <i>Streptomyces alni</i> , 67/53
orf01174	520	long-chain acyl-CoA synthetase	WP_047015640.1, <i>Streptomyces</i> sp. CNQ- 509, 82/74
orf01175	509	Puromycin resistance protein pur8	AKH84805.1, <i>Streptomyces</i> sp. CNQ-509, 80/70
orf01176	588	FAD-dependent oxidoreductases	WP_073501787.1, <i>Streptomyces paucisporeus</i> , 87/79

<https://doi.org/10.1371/journal.pone.0248404.t003>

The major interest of *Streptomyces* is its potential to produce diverse secondary metabolites with biological activities. Here, analysis using antiSMASH showed 27 other gene clusters in the genome of *S. niveus* SCSIO 3406 (Table 4). Of these gene clusters, some have the really low similarity with the known clusters, revealing the potential of *S. niveus* SCSIO 3406 to produce novel natural products. In the putative T1pks-oligosaccharide gene cluster (Table 4, cluster\_21), about 32% gene coding products showed similarity with the homologues of the known biosynthetic gene cluster of angucycline antibiotic, grincamycin, in *Streptomyces lusitanus* SCSIO LR32 [21]. We also identify an ectoine gene cluster (Table 4, cluster\_23) which

**Fig 3. Organization of the gene cluster of compounds 1–6 in *S. niveus* SCSIO 3406.**

<https://doi.org/10.1371/journal.pone.0248404.g003>

**Table 4. Putative gene cluster coding for secondary metabolites in *S. niveus* SCSIO 3406 via antiSMASH.**

No.	Type	From (bp)	To (bp)	Most similar known biosynthesis cluster	Similarity
1	NRPS-T1pks	16717	104129	Livipeptin	100
2	Other	184608	228006	–	–
3	NRPS-Linaridin	239107	354223	Laspartomycin	11
4	Terpene-otherks-T1pks	425830	492314	Carotenoid	54
5	T3pks	515177	556304	Herboxidiene	2
6	Terpene-otherks-lantipeptide	629756	732288	Coelichelin	100
7	Terpene	1046670	1073438	Hopene	69
8	terpene-Bacteriocin	1148109	1182287	–	–
9	Terpene	1439858	1461219	–	–
10	Bacteriocin	1563157	1574539	–	–
11	Terpene	1734097	1755776	–	–
12	butyrolactone	1763284	1774216	–	–
13	Siderophore	1883945	1898683	–	–
14	Other	2284811	2326703	Stambomycin	12
15	Siderophore	3740889	3756085	Macrotetrolide	33
16	Terpene	4086055	4109186	–	–
17	Terpene	4654606	4675796	Merochlorin	7
18	Bacteriocin	4774816	4785109	–	–
19	thiopeptide-lantipeptide	4856826	4891992	–	–
20	butyrolactone	5081056	5092780	–	–
21	T1pks-oligosaccharide	5780811	5848626	Grincamycin	32
22	Nrps	6228545	6274913	–	–
23	Ectoine	6339095	6349493	Ectoine	75
24	NRPS-T1pks	6640764	6720100	Streptomycin	19
25	otherks-Nrps	6975232	7045695	Naphthyridinomycin	92
26	T3pks	7312845	7353951	Alkylresorcinol	100
27	Melanin	7672825	7683463	Melanin	40

<https://doi.org/10.1371/journal.pone.0248404.t004>

consists of hydroxylase (Orf05563), L-ectoine synthase (Orf05564), diamino-butyr-2-oxo-glutarate transaminase (Orf05565) and L-2,4-diamino-butyr-4-oxo-glutarate acetyltransferase (Orf05566) in *S. niveus* SCSIO 3406; these gene coding products shows 75% similarity with the homologues of the known ectoine biosynthetic cluster (BGC0000853\_c1). As one kind of compatible solute, ectoine can be used for protecting enzymes, membranes and whole cells against stresses [22].

CRISPR (Clustered regularly interspaced short palindromic repeat) acronym was proposed by Jansen et al. [23]. CRISPR was observed first in 1987 in *Escherichia coli* [24] and were subsequently reported in a wide range of prokaryotic genomes. CRISPR associated proteins (Cas) use the CRISPR spacers to recognize and cut foreign genetic elements [25]. Therefore, the CRISPR/Cas system is a prokaryotic immune system [26]. Here, 22 CRISPRs candidates spreading over *S. niveus* SCSIO 3406 genome, including 6 confirmed CRISPRs and 16 questionable CRISPRs [27], were identified in *S. niveus* SCSIO 3406 genome via CRISPRFinder, well above the average level in *Streptomyces* whose genome sequences have been published (Table 5). The CRISPR sequence contains 168 spacer sequences in size from 19 to 99 bp. The number of spacers in each locus varies from 1 to 78. Only two of the spacer sequences (5' -g cgcgacggacgcgcccgggtgagcacgcgcaggg-3' and 5' -gtcctcggtccggttc gtcctgcgcgatctccag-3'), namely, protospacers match any sequences in the public sequence databases. The rest of the spacers remain the CRISPR “dark matter”. We also

Table 5. Number of CRISPRs in *Streptomyces* genome.

Strains	Number of confirmed CRISPRs	Number of questionable CRISPRs	Total
<i>Streptomyces albus</i> J1074	0	0	0
<i>Streptomyces avermitilis</i> MA-4680	3	4	7
<i>Streptomyces bingchenggensis</i> BCW-1	3	7	10
<i>Streptomyces cattleya</i> NRRL 8057	2	8	10
<i>Streptomyces coelicolor</i> A3(2)	1	2	3
<i>Streptomyces collinus</i> Tu 365	0	4	4
<i>Streptomyces davawensis</i> JCM 4913	0	0	0
<i>Streptomyces flavogriseus</i> ATCC 33331	4	2	6
<i>Streptomyces fulvissimus</i> DSM 40593	0	0	0
<i>Streptomyces griseus</i> subsp. <i>griseus</i> NBRC 13350	3	7	10
<i>Streptomyces hygroscopicus</i> subsp. <i>jinggangensis</i> 5008	9	4	13
<i>Streptomyces hygroscopicus</i> subsp. <i>jinggangensis</i> TL01	0	0	0
<i>Streptomyces rapamycinicus</i> NRRL 5491	10	4	14
<i>Streptomyces scabiei</i> 87.22	0	2	2
<i>Streptomyces</i> sp. PAMC26508	0	0	0
<i>Streptomyces</i> sp. SirexAA-E	0	4	4
<i>Streptomyces venezuelae</i> ATCC 10712	0	6	6
<i>Streptomyces violaceusniger</i> Tu 4113	5	12	17
<i>S. niveus</i> SCSIO 3406	6	16	22

<https://doi.org/10.1371/journal.pone.0248404.t005>

identified ten genes (*orf03151*, *orf05316*, *orf06507-orf06514*) coding for CRISPR-associated proteins. CRISPR loci together with *cas* (CRISPR-associated) genes form the powerful immune system for *S. niveus* SCSIO 3406.

Interestingly, several antibiotic resistance genes were identified in *S. niveus* SCSIO 3406 genome. The gene *orf04161* encodes penicillin amidase catalyzing the hydrolysis of benzylpenicillin [28], that efficiently accounts for the fact *S. niveus* SCSIO 3406 can survive on plates which contained 100µg/ml penicillin; the gene *orf00505* encodes erythromycin esterase hydrolyzing the lactone ring of the 14 membered macrolides erythromycin and oleandomycin [29]; the gene *orf02495* encodes virginiamycin B lyase inactivating the type B streptogramin antibiotics by linearizing the lactone ring at the ester linkage [30].

In summary, we have completely sequenced the genome of the deep South China Sea-derived *S. niveus* SCSIO 3406. By bioinformatic analysis, we have identified a biosynthetic gene cluster for both the marfuraquinocins A-D and phenaziterpenes A-B. The identified gene cluster provides important genetic basis for better understanding of biosynthesis mechanism of the marfuraquinocins A-D and phenaziterpenes A-B, as well as the construction of improved strain for the industrial production. Notably, 27 other gene clusters were also predicted in the genome of *S. niveus* SCSIO 3406. More importantly, some of these clusters have the really low similarity with the known clusters, strongly suggesting the potential of *S. niveus* SCSIO 3406 to produce diversity of novel natural products. This sequence information paves the way for the genome mining of *S. niveus* SCSIO 3406 for the novel natural product discovery.

## Author Contributions

**Conceptualization:** Jianhua Ju, Qinglian Li.

**Data curation:** Qinghua Zhu, Qinglian Li.



**Formal analysis:** Qinghua Zhu, Weige Cheng, Yongxiang Song, Qing He, Jianhua Ju, Qinglian Li.

**Funding acquisition:** Qinghua Zhu, Yongxiang Song, Qinglian Li.

**Project administration:** Jianhua Ju.

**Resources:** Yongxiang Song.

**Supervision:** Jianhua Ju.

**Writing – original draft:** Qinghua Zhu, Qinglian Li.

**Writing – review & editing:** Qing He, Qinglian Li.

## References

1. Sivalingam P, Hong K, Pote J, Prabakar K. Extreme environment *Streptomyces*: potential sources for new antibacterial and anticancer drug leads? *Int J Microbiol.* 2019; 2019: 5283948. <https://doi.org/10.1155/2019/5283948> PMID: 31354829
2. Fenical W, Jensen PR. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat Chem Biol.* 2006; 2(12): 666–73. <https://doi.org/10.1038/nchembio841> PMID: 17108984
3. Pathirana C, Jensen PR, Fenical W. Marinone and debromomarinone: Antibiotic sesquiterpenoid naphthoquinones of a new structure class from a marine bacterium. *Tetrahedron Lett.* 1992; 33: 7663–6. [https://doi.org/10.1016/0040-4039\(93\)88010-G](https://doi.org/10.1016/0040-4039(93)88010-G)
4. Cho JY, Kwon HC, Williams PG, Jensen PR, Fenical W. Azamerone, a terpenoid phthalazinone from a marine-derived bacterium related to the genus *Streptomyces* (Actinomycetales). *Org Lett.* 2006; 8(12): 2471–4. <https://doi.org/10.1021/ol060630r> PMID: 16737291
5. Soria-Mercado IE, Prieto-Davo A, Jensen PR, Fenical W. Antibiotic terpenoid chloro-dihydroquinones from a new marine actinomycete. *J Nat Prod.* 2005; 68(6): 904–10. <https://doi.org/10.1021/np058011z> PMID: 15974616
6. Leipoldt F, Zeyhle P, Kulik A, Kalinowski J, Heide L, Kaysser L. Diversity of ABBA prenyltransferases in marine *Streptomyces* sp. CNQ-509: Promiscuous enzymes for the biosynthesis of mixed terpenoid compounds. *PLoS One.* 2015; 10(12): e0143237. <https://doi.org/10.1371/journal.pone.0143237> PMID: 26659564
7. Song Y, Huang H, Chen Y, Ding J, Zhang Y, Sun A, et al. Cytotoxic and antibacterial marfuraquinocins from the deep South China Sea-derived *Streptomyces niveus* SCSIO 3406. *J Nat Prod.* 2013; 76(12): 2263–8. <https://doi.org/10.1021/np4006025> PMID: 24251399
8. Qin S, Chen HH, Zhao GZ, Li J, Zhu WY, Xu LH, et al. Abundant and diverse endophytic actinobacteria associated with medicinal plant *Maytenus austroyunnanensis* in Xishuangbanna tropical rainforest revealed by culture-dependent and culture-independent methods. *Environ Microbiol Rep.* 2012; 4(5): 522–31. <https://doi.org/10.1111/j.1758-2229.2012.00357.x> PMID: 23760897
9. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics.* 2008; 24(5): 713–4. <https://doi.org/10.1093/bioinformatics/btn025> PMID: 18227114
10. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, et al. A whole-genome assembly of *Drosophila*. *Science.* 2000; 287(5461): 2196–204. <https://doi.org/10.1126/science.287.5461.2196> PMID: 10731133
11. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics.* 2007; 23(6): 673–9. <https://doi.org/10.1093/bioinformatics/btm009> PMID: 17237039
12. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics.* 2011; 12: 444. <https://doi.org/10.1186/1471-2164-12-444> PMID: 21899761
13. 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(Web Server issue): W339–46. <https://doi.org/10.1093/nar/gkr466> PMID: 21672958
14. Flinspach K, Rückert C, Kalinowski J, Heide L, Apel AK. Draft Genome Sequence of *Streptomyces niveus* NCIMB 11891, Producer of the Aminocoumarin Antibiotic Novobiocin. *Genome Announc.* 2014; 2(1): e01146–13. <https://doi.org/10.1128/genomeA.01146-13> PMID: 24407644
15. Steinbüchel A. Production of rubber-like polymers by microorganisms. *Curr Opin Microbiol.* 2003; 6(3): 261–70. [https://doi.org/10.1016/s1369-5274\(03\)00061-4](https://doi.org/10.1016/s1369-5274(03)00061-4) PMID: 12831902

16. Cornforth RH, Popják G, Yengoyan L. Studies on the biosynthesis of cholesterol. XX. Steric course of decarboxylation of 5-pyrophosphomevalonate and of the carbon to carbon bond formation in the biosynthesis of farnesyl pyrophosphate. *J Biol Chem.* 1966; 241(17): 3970–87. PMID: [4288360](https://pubmed.ncbi.nlm.nih.gov/4288360/)
17. Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S. A new pathway for polyketide synthesis in microorganisms. *Nature.* 1999; 400(6747): 897–9. <https://doi.org/10.1038/23748> PMID: [10476972](https://pubmed.ncbi.nlm.nih.gov/10476972/)
18. Martin JF, Liras P. Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu Rev Microbiol.* 1989; 43: 173–206. <https://doi.org/10.1146/annurev.mi.43.100189.001133> PMID: [2679353](https://pubmed.ncbi.nlm.nih.gov/2679353/)
19. Takahashi S, Koyama T. Structure and function of *cis*-prenyl chain elongating enzymes. *Chem Rec.* 2006; 6(4): 194–205. <https://doi.org/10.1002/tcr.20083>
20. Cornforth JW, Fujisaki S, Hara H, Nishimura Y, Horiuchi K, Nishino T. Cloning and nucleotide sequence of the *ispA* gene responsible for farnesyl diphosphate synthase activity in *Escherichia coli*. *J Biochem.* 1990; 108(6): 995–1000. <https://doi.org/10.1093/oxfordjournals.jbchem.a123327> PMID: [2089044](https://pubmed.ncbi.nlm.nih.gov/2089044/)
21. Zhang Y, Huang H, Chen Q, Luo M, Sun A, Song Y, et al. Identification of the grincamycin gene cluster unveils divergent roles for GcnQ in different hosts, tailoring the L-rhodinose moiety. *Org Lett.* 2013; 15(13): 3254–7. <https://doi.org/10.1021/ol401253p> PMID: [23782455](https://pubmed.ncbi.nlm.nih.gov/23782455/)
22. Lippert K, Galinski EA. Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. *Appl Microbiol Biotechnol.* 1992; 37: 61–5. <https://doi.org/10.1007/BF00174204>
23. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.* 2002; 43(6): 1565–75. <https://doi.org/10.1046/j.1365-2958.2002.02839.x> PMID: [11952905](https://pubmed.ncbi.nlm.nih.gov/11952905/)
24. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol.* 1987; 169(12): 5429–33. <https://doi.org/10.1128/jb.169.12.5429-5433.1987> PMID: [3316184](https://pubmed.ncbi.nlm.nih.gov/3316184/)
25. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007; 315(5819): 1709–12. <https://doi.org/10.1126/science.1138140> PMID: [17379808](https://pubmed.ncbi.nlm.nih.gov/17379808/)
26. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet.* 2010; 11(3): 181–90. <https://doi.org/10.1038/nrg2749> PMID: [20125085](https://pubmed.ncbi.nlm.nih.gov/20125085/)
27. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics.* 2007; 8: 172. <https://doi.org/10.1186/1471-2105-8-172> PMID: [17521438](https://pubmed.ncbi.nlm.nih.gov/17521438/)
28. Buchholz K. A breakthrough in enzyme technology to fight penicillin resistance-industrial application of penicillin amidase. *Appl Microbiol Biotechnol.* 2016; 100(9): 3825–39. <https://doi.org/10.1007/s00253-016-7399-6> PMID: [26960323](https://pubmed.ncbi.nlm.nih.gov/26960323/)
29. Barthélémy P, Autissier D, Gerbaud G, Courvalin P. Enzymic hydrolysis of erythromycin by a strain of *Escherichia coli*. A new mechanism of resistance. *J Antibiot (Tokyo).* 1984; 37(12): 1692–6. <https://doi.org/10.7164/antibiotics.37.1692> PMID: [6396291](https://pubmed.ncbi.nlm.nih.gov/6396291/)
30. Mukhtar TA, Koteva KP, Hughes DW, Wright GD. Vgb from *Staphylococcus aureus* inactivates streptogramin B antibiotics by an elimination mechanism not hydrolysis. *Biochemistry.* 2001; 40(30): 8877–86. <https://doi.org/10.1021/bi0106787> PMID: [11467949](https://pubmed.ncbi.nlm.nih.gov/11467949/)